

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

---

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**





No.	Score	Match	Length	DB	ID	Description
C 1	578.2	25.9	616	61	AL045824	AL045824 DKFPA34H
C 2	477.2	21.3	512	79	AM300030	AM300030 xs55h01.x
C 3	457.8	20.5	473	44	AI269960	AI269960 qk18d11.x
C 4	440.4	19.7	485	39	AA660926	AA660926 ak22d06.s
C 5	437.6	19.6	467	51	AI761354	AI761354 w16e08.x
C 6	435.4	19.5	464	51	AI739238	AI739238 w16e07.x
C 7	423	18.9	473	63	AI984995	AI984995 wr89g06.x
C 8	405.6	18.1	636	26	W22160	W22160 63A6 Human
C 9	403	18.0	411	46	AI628985	AI628985 cy79c02.x
C 10	403	18.0	411	49	AI629007	AI629007 ty79e02.x
C 11	394.2	17.6	407	63	AI989344	AI989344 w27e03.x
C 12	372.8	16.7	384	60	AI796741	AI796741 wai4d02.x
C 13	371.2	16.6	385	44	AI722081	AI722081 qj78f01.x
C 14	371.2	16.6	465	80	AM390711	AM390711 RC1-ST019
C 15	369.2	16.5	386	46	AI453367	AI453367 t3j7c11.x
C 16	355.2	15.9	360	63	AI991419	AI991419 ws10b04.x
C 17	351.4	15.7	367	79	AM304132	AM304132 xs13f01.x
C 18	341.4	15.3	381	80	AM351654	AM351654 QV2-CT014
C 19	335.6	15.0	342	51	AI738967	AI738967 w12f02.x
C 20	334	14.9	342	51	AI738958	AI738958 w12e03.x
C 21	304.6	13.6	311	35	AA551799	AA551799 nk04a11.s
C 22	292.2	13.1	317	32	AA344913	AA344913 EST50856
C 23	290.4	13.0	335	28	AA121174	AA121174 z188g08.s
C 24	287.8	12.9	316	74	AM192301	AM192301 x184a12.x
C 25	270.4	12.1	272	44	AI270726	AI270726 q193d07.x
C 26	267.4	12.0	289	63	AA348243	AA348243 EST54707
C 27	266.4	11.9	278	63	AM013904	AM013904 UI-H-B10
C 28	248.8	11.1	279	61	AI869919	AI869919 we75h03.x
C 29	240.8	10.8	244	60	AI783811	AI783811 lu79c11.x
C 30	197.4	8.8	216	61	AI869907	AI869907 we75g03.x
C 31	180.8	8.1	206	31	AA327309	AA327309 EST30621
C 32	160	7.2	204	28	AA121198	AA121198 z188g08.r
C 33	143	6.4	353	49	AI642234	AI642234 vn8e05.y
C 34	131.2	5.9	221	64	AM062449	AM062449 MR0-CT006
C 35	120.6	5.4	473	36	AA616807	AA616807 vn8e05.r
C 36	105	4.7	463	105	AQ320634	AQ320634 RPT11-92
C 37	68.2	3.1	823	84	B03396	B03396 cSRL-17BD10
C 38	67.2	3.0	443	102	AQ100902	AQ100902 HS-3060_B
C 39	61.6	2.8	997	82	CNS005TE	AL060767 Drosophila
C 40	57.2	2.6	587	36	C79507	C79507 Mous
C 41	56.6	2.5	571	92	AQ919110	AQ919110 RPT1-23-2
C 42	55.6	2.5	476	36	C79500	C79500 Mous
C 43	55.6	2.5	445	34	AA646787	AA646787 zxb8g10.r
C 44	55.2	2.5	498	50	AI676456	AI676456 etestp021
C 45	54.6	2.4	657	80	AM369742	AM369742 IL0-BT015

## ALIGNMENTS

RESULT 1  
AL045824/c  
LOCUS AL045824 616 bp mRNA EST 29-SEP-1999  
DEFINITION DKFPA34H336\_s1 434 (synonym: htes3) Homo sapiens CDNA clone  
ACCESSION DKFPA34H336 3', mRNA sequence.  
VERSION AL045824  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 616)  
AUTHORS Koehner, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.  
TITLE EST (Koehner, et al.)  
JOURNAL Unpublished (1999)  
COMMENT On May 18, 1998 this sequence version replaced gi:3137188.  
Contact: Koehner K  
MIS  
Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 3' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
Sequenced by BMF within the CDNA sequencing consortium of the German Genome Project.  
ri sequence also available.  
This clone is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers

## FEATURES

Source

1. 616  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFPA34H336"  
/clone\_lib="434 (synonym: htes3)"  
/tissue\_type="testis"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pSPort1; Site\_1: NotI; Site\_2: SalI"

BASE COUNT 134 a 142 c 185 g 155 t

## ORIGIN

Query Match 25.9%; Score 578.2; DB 61; Length 616;  
Best Local Similarity 98.4%; Pred. No. 1.3e-134;  
Matches 605; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

QY 1603 atggcgaataaaggaaagttgtgaagactaaccttcgtgaggtacgtgcctg 1662  
DB 613 atggcgaataaaggaaagttgtgaagactaaccttcgtgaggtacgtgcctg 554  
QY 1663 cagaagccttgtaatacgaagcgagactgacgaactgtgttcgctcgtcccca 1722  
DB 553 cagaagccttgtaatacgaagcgagactgacgaactgtgttcgctcgtcccca 494  
QY 1723 aaggagatgacgaatgagctgacacttctgtcgaagtcacaagggaaccaatcac 1782  
DB 493 aaggagatgacgaatgagctgacacttctgtcgaagtcacaagggaaccaatcac 435  
QY 1783 gggtagaagaacagtagctcggcccccacctcgtgaacacaaagctcggagccct 1842  
DB 434 -cgttagaagaacagtagctcggcccccacctcgtgaacacaaagctcggagccct 376  
QY 1843 gaggagcgacatgacacacacacagatgacatccttggagactgaatcacccag 1902  
DB 375 gaggagcgacatgacacacacacagatgacatccttggagactgaatcacccag 316  
QY 1903 aatcacaactcccttggccctgaacgaagccagataaggaacagctcggacactt 1962  
DB 315 aatcacaactcccttggccctgaacgaagccagataaggaacagctcggacactt 256  
QY 1963 tttagaagcacaatgtggaagaaaggaagcagccgtttggaagaaatcagaagt 2022  
DB 255 tttagaagcacaatgtggaagaaaggaagcagccgtttggaagaaatcagaagt 196  
QY 2023 ccagactccatccttcccttgcctcgtgcccagtgatgtgtcctccagcttggggag 2082  
DB 195 ccagactccatccttcccttgcctcgtgcccagtgatgtgtcctccagcttggggag 136  
QY 2083 tccctccttgaacccataaagaacccactggaagctcctctctccatccctcctc 2142  
DB 135 tccctccttgaacccataaagaacccactggaagctcctctctccatccctcctc 76  
QY 2143 gccctcgtcctaattgctgccaagatgtgaactcaaaccttactcagctcctaat 2202  
DB 75 gccctcgtcctaattgctgccaagatgtgaactcaaaccttactcagctcctaat 16  
QY 2203 aaataaacaagatt 2217  
DB 15 aaataaacaagatt 1

RESULT 2  
AM300030/c



LOCUS AM300030 512 bp mRNA EST 18-JAN-2000  
 DEFINITION xs5501.x1 NCI\_CGAP\_kid11 Homo sapiens cdna clone IMAGE:2773585 3',  
 mRNA sequence.  
 ACCESSION AM300030  
 VERSION AM300030.1 GI:6709707  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 512)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 TITLE Unpublished (1997)  
 JOURNAL On May 18, 1998 this sequence version replaced gi:3137941.  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.11nl.gov/bdrip/image/image.html](http://www.bio.11nl.gov/bdrip/image/image.html)

Seq primer: -400P from Gibco  
 High quality sequence stop: 462.  
 Location/Qualifiers  
 1..512  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2773585"  
 /clone\_1lb="NCI\_CGAP\_kid11"  
 /lab\_host="DH10B"  
 /note="Organ: kidney; Vector: pRT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
 Plasmid DNA from the normalized library NCI\_CGAP\_kid3 was  
 prepared, and ss circles were made in vitro. Following HAP  
 purification, this DNA was used as tracer in a subtractive  
 hybridization reaction. The driver was PCR-amplified cDNAs  
 from a pool of 5,000 clones made from the same library  
 (cloneids 1322376-1323911, 1456007-1456775, and  
 1500552-1502855). Subtraction by Bento Soares and M.  
 Fatima Bonaldo."  
 Fatima Bonaldo.

BASE COUNT 117 a 105 c 158 g 132 t  
 ORIGIN

Query Match 21.3%; Score 477.2; DB 79; Length 512;  
 Best Local Similarity 98.1%; Pred. No. 2.7e-109;  
 Matches 504; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

QY 1712 tctgcccccaagagatgaagatgaagcttcttgcgtgaagtccaaggaggaa 1771  
 |||||||  
 DB 512 TCTGCCCCCAAGAGATGACATGAGCTGACCTCTT-CTGAGTCCAAAAGGGGAA 454  
 |||||||

QY 1772 accaatcaagcgcttggaagacagtagctcgcccaacctctgaacacaagcctc 1831  
 |||||||  
 DB 453 ACCAAATTCAC-CCTTGTGGAACAGTAGACTCCGGCCCACTGTGTAACACAAAGCCTC 395  
 |||||||

QY 1832 ggaacacgcttgagagagccacatgacacacacagatgcatccttggagcctgaatc 1891  
 |||||||  
 DB 394 GGATCACCCTTGAAGAGAGCCACACTACACACACAGATGGCANTCTTGGGACCTGAATC 335  
 |||||||

QY 1892 tatcacccagaatctcaaacctcttgcctgaaccagggccagataaggaacagct 1951  
 |||||||  
 DB 334 TATCACCACAGGAATCTCAAACTCCCTTGGCCCTGAACCAAGGGCAGATGAAGAACAGCT 275  
 |||||||

QY 1952 cgggcccattttgaaggccaatgtgagagaaaggagagcagccgctttggagaag 2011  
 |||||||

DB 274 CGGGCACACTTCTTGAAAGGCCAACGTGGAGGAAGGAGCAGCCAGCATTTGGGAGAG 215  
 QY 2012 atccaagagatccaagactctatcctctctctctgcgcagtgatgtctctccag 2071  
 |||||||  
 DB 214 ATCTCAAGGATCCAGACTCTCATTTCTCTCGGCGCACTGGAATTTGGCTCTCCAG 155  
 |||||||

QY 2072 cttgggggaactctctcttgaaccctaataagaccacagtgagctctctctcat 2131  
 |||||||  
 DB 154 CTCTGGGGGACTCTTCTCTTGAACCTTAATTAAGACCCCACTGGAGTCTCTCTCAT 95  
 |||||||

QY 2132 cccctctctgcgcctctgtcttaattgctgcagagattgtcactccaactactctg 2191  
 |||||||  
 DB 94 CCTCTCTCTGCGCCCTGCTCTTAATTCGTCGCGAGATTTGTCACCTCAACCTTACTG 35  
 |||||||

QY 2192 agctcattataaataacagattatttcca 2225  
 |||||||  
 DB 34 AGCTCATTAATTAATTAACAGATTTATTTTCCA 1  
 |||||||

RESULT 3  
 AI269960/c 473 bp mRNA EST 28-JAN-1999  
 LOCUS qk18d11.x1 NCI\_CGAP\_kid3 Homo sapiens cdna clone IMAGE:1869333 3'  
 DEFINITION similar to TR:Q14160 Q14160 KIA0147 PROTEIN ;, mRNA sequence.  
 ACCESSION AI269960  
 VERSION AI269960.1 GI:3889127  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 473)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.11nl.gov/bdrip/image/image.html](http://www.bio.11nl.gov/bdrip/image/image.html)

Insert Length: 520 Std Error: 0.00  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 451.  
 Location/Qualifiers  
 1..473  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1869333"  
 /clone\_1lb="NCI\_CGAP\_kid3"  
 /lab\_host="DH10B"  
 /note="Organ: kidney; Vector: pRT73D-Pac (Pharmacia) with  
 a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer,  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not  
 I and Eco RI sites of the modified pRT73 vector. mRNA  
 source: 2 pooled kidneys. Library went through one round  
 of normalization. Library constructed by Bento Soares and  
 M. Fatima Bonaldo."

BASE COUNT 92 a 159 c 118 g 104 t  
 ORIGIN

Query Match 20.5%; Score 457.8; DB 44; Length 473;  
 Best Local Similarity 99.6%; Pred. No. 1.9e-104;

Matches 459; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 654 tcagttgtgtcgaatcgtcgggagctgcgaagcagccttgccctcgaatcggga 713
Db 461 TCAGTTGTGTGGAATCTGGGGGCTGGAGGACACCTGGGCTCCCTTGGAATCGGA 402
QY 714 aaacagaagaagaagcttcacacagcctgtgtagctcccgagccttgctcagcat 773
Db 401 AAACAAGAAGAAGAGTCTTCATCAGCCTGGTAGGCTCCCGAGGCTTGCTGCAGCAT 342
QY 774 ttccagggcccccacacagaagcctgtgcatcttaccagcattgtgaacgtgctccct 833
Db 341 TTCACAGGCGCCCATCAAGCCTTGCACTTTATCAGCATGTGAACCTGGCTCCCT 282
QY 834 gctcgtcgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 893
Db 281 GCTCCTGTGAGGTGGAGTTGGAGATAGGGACAGATTGTGAGTGAATGGCTGCACCTT 222
QY 894 ccttaacctgtatcacaagaaggctgtlaaatgtcgtgaataatagccgagcctgacat 953
Db 221 CCTTAACCTGTGATCACAAGAGGCTGTAAATGTGCTGAAGAGTAGCCGACGCTGACAT 162
QY 954 ctccattgtagctgcagctgtgcggagagctgttcatgacaagcgggagcgctgtgcaga 1013
Db 161 CTCCTATTGTAGCTGCAGCTGTGGCGGAGCGTTCATGACAGACCGGAGCGGCTGTGCAGA 102
QY 1014 ggcggcgagcgcgtgagctcagcgagcagcgtcttcacatcagcagaagcggctgagatga 1073
Db 101 GCGCGCGGACGCTGAGCTGACGCGGAGGAGCTTCTCATGCAAGACGCGCTGGCGATTGA 42
QY 1074 gtccacaagaatcctcccaaggagcagcagcagcagcagcagcagcagcagcagcagc 1114
Db 41 GTCCACAAGATCTCTCCAGAGCAGCAGAGATGAGAGCGGC 1

```

RESULT 4  
LOCUS AA860926/c 485 bp mRNA EST 04-JAN-1999  
DEFINITION ak22406.s1 Soares testis\_NHT Homo sapiens cDNA clone IMAGE:1406699  
ACCESSION AA860926  
VERSION AA860926.1 GI:2953066  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 485)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL On Jan 19, 1998 this sequence version replaced gi:2287585.  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: Robert.Strausberg@nih.gov  
Tel: (301) 496-1550  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/MLNL at:  
www.bio.lnl.gov/bbrp/image/image.html

FEATURES  
Source  
1. 485  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1406699"  
/clone\_lib="Soares\_testis\_NHT"  
Location/Qualifiers  
Insert Length: 629 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence scop: 426.

/sex="male"  
/lab\_host="DH10B"  
/note="Vector: pUT3D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech  
Laboratories, Inc., and primed with a Not I - oligo(dT)  
primer [5',  
TGTTCACATCTGAGAGTGAGAGCGGCGCCCAATTTTTTTTTTTT 3'].  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pUT3D vector. Library  
went through one round of normalization to Col5, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

Query Match 19.7%; Score 440.4; DB 39; Length 485;  
Best Local Similarity 98.1%; Pred. No. 4.5e-100;  
Matches 477; Conservative 0; Mismatches 6; Indels 3; Gaps 3;

```

QY 1740 gctgacctctgtcgtgaagtcacaaagggaacacaaatcacggttaggaacagtga 1799
Db 485 GCTGACCTTCTT-CTGAAGTCCAAAAGGGGAAACCAAAATTGAC-CGTTAGGAAACAGTGA 428
QY 1800 gctcggcccaacctcgtgtaacaaagaagcctcggaacagccttgagaagagccacatgac 1859
Db 427 GCTCGGCGCCCACTCTGTGAACAAAGCCTCGGACACCTTGAGAGAGGCCACACTAC 368
QY 1860 aaacacagaatgacatccttggagcctgaatcatalcaccagaagatcacaatccctt 1919
Db 367 ACACACCAATGAGCATCTTGGGACCTGATCTATCACCAGAGAAATCTCAAACTCCCTTT 308
QY 1920 ggcgcctgaacacagggtcaataagaaagcgtcgggacacttttgaaggccaatgag 1979
Db 307 AG-CCTGAACCAAGGGCCAGATAGGAACAGCTGGGCGCACCTTCTGAAGGCCAATGTGG 249
QY 1980 aggaagaaggagcagcagcagcgtttggaagaagatcgaagatccagactccatcctt 2039
Db 248 AGGAAGAAGGAGCAGCAGCGCTTTGGGAGAAAGATCTCAAGATCTCATTTCTT 189
QY 2040 tccctcgtccagtgaaattgctcctcccaagccttggggagcctcctcctgaaccccta 2099
Db 188 TCCTCTGGCCCAAGTAATTGGTCTCTCCACACTCTGGGGAGCTCCTTCTGAACCCCTTA 129
QY 2100 ataagaacccacatgagctcctcctcctccatccctcctcctcctcctcctcctcct 2159
Db 128 ATAGAAGCCCACTGAGTCTCTCTCTCCATCCCTCTCTGCGCTGCTTAATTG 69
QY 2160 ctgcagagattgtcactccaacacttactctgagctcattataaataacagattat 2219
Db 68 CTGCCAGGATTGTCACTCCAAACCTTACTCTGAGCTATTATAAATAACAGATTAT 9
QY 2220 ttccca 2225
Db 8 TTTCGA 3

```

RESULT 5  
LOCUS A1761354/c 467 bp mRNA EST 24-JUN-1999  
DEFINITION w160e08.x1 NCI-CGAP\_Col6 Homo sapiens cDNA clone IMAGE:2394662 3',  
RNA sequence.  
ACCESSION A1761354  
VERSION A1761354.1 GI:5177021  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 467)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL  
COMMENT

Tumor Gene Index  
Unpublished (1997)  
On May 18, 1998 this sequence version replaced gi:3138298.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
www-bio.1nl.gov/bdrp/image/image.html

FEATURES  
Source

Seq primer: -400P from GIBCO.  
Location/Qualifiers  
1. 467  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NCI-CGAP-Co16"  
/tissue\_type="colon tumor, RER+"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; plasmid DNA from the normalized library NCI-CGAP-Co16 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1057416-1061255, and 1144584-1145351).  
Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 108 a 96 c 143 g 120 t

ORIGIN

Query Match 19.6%; Score 437.6; DB 51; Length 467;

Best Local Similarity 98.7%; Pred. No. 2.2e-99; Mismatches 4; Indels 2; Gaps 2;

Matches 462; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 1765 aggggaaacaaatcagcgttagaagagtagctccgcccacacccctgtaacaca 1824  
|||||  
DB 467 AGGGGAACCAATTCAC-CGTTAGAAACAGTAGCTCCGGCCCACTCGTGAACACA 409  
|||||  
QY 1825 aagcctcgagcagccttgagagagccacatgacacacacagatgcatccttggagc 1884  
|||||  
DB 408 AAGCCTCGACACAGCTTGAAGAGAGCCACACTACACACACAGATGGCATCTTGGGAC 349  
|||||  
QY 1885 ctgaatctatcacccaggaatctcaactcctcttgccctggaacccagggccagataag 1944  
|||||  
DB 348 CTGAATCTATCACCAAGAAATCTCAACCTCTTGGCCCTGGAACCAAGGCAATTAAG 289  
|||||  
QY 1945 aacagctcgagccact-ttttgaagccaatgtgaggaaggaagcagccagcgtt 2003  
|||||  
DB 288 AACAGCTCGGGCCACTCTTTTGAAGCCATGTGAGAAAGGACACAGCCGTTT 229  
|||||  
QY 2004 gggaggaagatcgaagatccagatctcatctcttccctctgagccagtgaaattgctc 2063  
|||||  
DB 228 GGGAGAGATCTCAAGGATCCAGACACTCTCATCTCTCTGCGCCAGTAATTTGGTC 169  
|||||  
QY 2064 tctccagccttgggggagcctctccttgaaacctataagaccacacagagctctctc 2123  
|||||  
DB 168 TCTCCAGCTCTGGGAGACTCTTCTTGAACCTTAATTAAGCCCACTGAGTCTCT 109  
|||||  
QY 2124 ctctccatccctcctcctcctcctctgctctaatctgctgccaagatgtcaactcaaac 2183  
|||||  
DB 108 CTCGCCATCCCTCTCCCTGCTCTGCTCTGAATTTGCTGCCAGATGTCTCACTCAAAAC 49  
|||||  
QY 2184 tctactgagctcaitaataaataacagattatttccagcttaa 2231  
|||||  
DB 48 TTACTGTAGCTCATTAATAAATAACAGATTATTTTCAGCTAAA 1

RESULT 6  
LOCUS A1739238/c  
DEFINITION w114407.x1 NCI-CGAP-Co16 Homo sapiens cDNA clone IMAGE:2390268 3', mRNA sequence.  
ACCESSION A1739238  
VERSION A1739238.1 GI:5101219  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 464)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
On Jun 5, 1998 this sequence version replaced gi:31389064.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
www-bio.1nl.gov/bdrp/image/image.html

JOURNAL

COMMENT

FEATURES  
Source

Insert length: 582 Std Error: 0.00  
Seq primer: -400P from GIBCO  
High quality sequence stop: 421.  
Location/Qualifiers  
1. 464  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NCI-CGAP-Co16"  
/tissue\_type="colon tumor, RER+"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; plasmid DNA from the normalized library NCI-CGAP-Co16 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1057416-1061255, and 1144584-1145351).  
Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 104 a 90 c 140 g 130 t

ORIGIN

Query Match 19.5%; Score 435.4; DB 51; Length 464;

Best Local Similarity 98.7%; Pred. No. 7.8e-99; Mismatches 6; Indels 0; Gaps 0;

Matches 439; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1792 aacgtgagctccggcccacacctctgtaacacaaagctcgaccagccttgagagagc 1851  
|||||  
DB 464 AACAGTAACTCCGGCCCACTCTGTAAACAAAGCTCGACCACTTGAAGAGGC 405  
|||||  
QY 1852 cacatgacacacacagatgcatccttggagcctgatatcaccacagaaatcctaa 1911  
|||||  
DB 404 CACACTACACACACAGATGGCATCTTGGAGCTGATGATACCCAGGAATCTCAAA 345  
|||||  
QY 1912 ctcccttggccttgagacccagggccagataaggaagcctcgagcacttttgaagc 1971  
|||||  
DB 344 CTCCTTTGGCCCTGGAACCAAGGCGCAGATGAAGAACGCTCGGCGCACTTCTGAAGGC 285  
|||||  
QY 1972 caatgtgaggaaggaagcagcagccttggagaaagatcctcaagagatccagactc 2031  
|||||

Db 284 CAATGTGAGGAAGGAGCAGCAGCCGTTTGGAGAGATCTCAAGATCCAGACTCT 225  
Qy 2032 ccttccttcctcctgagccaggaatttgctcctccagcttgaggagcttcctt 2091  
Db 224 CATTCCTTCTCCTGAGCCAGTAATTTGGTCTCTCCAGCTTGAGGAGACTCTCTT 165  
Qy 2092 gaaccctaaagaagccacactgagctctctctcctccatccctcctctcctctg 2151  
Db 164 GAACCTTAATAGACCCAGCTGAGTCTCTCTCCATCCCTCTCCTGCGCTCTG 105  
Qy 2152 tctaatgtgcagagatgtcaactcaaaccttactctgagctcttataataaac 2211  
Db 104 TCTAATTTGGTCCAGGATTTGTCATCTCAAACTTACTGTGAGCTCATTAATAAATAAC 45  
Qy 2212 agattatttccagcttaaaaaa 2236  
Db 44 AGATTATTCTCCAGCTAAAAAAA 20

RESULT 7  
A1984995/c 473 bp mRNA EST 31-AUG-1999  
LOCUS w89g06.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2494906 3',  
DEFINITION mRNA sequence.  
ACCESSION A1984995  
VERSION A1984995.1 GI:5812272  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 473)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 18, 1998 this sequence version replaced gi:3138171.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. consortium/ILN at:  
www.bio.lnlnl.gov/bdrp/image/image.html

FEATURES  
source  
Seq primer: -40UP from Gibco  
High quality sequence stop: 463.  
Location/Qualifiers  
1..473  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2494906"  
/clone\_lib="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker. Site\_1: Not 1; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(clonoids 132376-132391, 145607-145675, and  
150052-150285). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 110 a 96 c 145 g 121 t 1 others  
ORIGIN

Query Match 18.9%; Score 423; DB 63; Length 473;

Best Local Similarity 96.0%; Pred. No. 1e-95;  
Matches 455; Conservative 0; Mismatches 16; Indels 3; Gaps 2;  
Qy 1751 tctgtgaagccaaaagggagaaacaaatcagcgtttaagaaacgttgctccggccc 1810  
Db 471 TTTCTGAAGTCCAAAAGGGGAAACCAAAATTCAC-CGTTAGGAAACAGTGAAGCTCGGCC 413  
Qy 1811 acctgtgaacacaaagctcgaccagccttgaagagagccacatgaacacacacagat 1870  
Db 412 ACCTGCTGAACACANAGCTCGGATC--AGCTTAGAGAGAGCCACACTACACACAGAT 355  
Qy 1871 ggcattccttggagacctgaatctataccacaggaatctcaactcccttggccctgaac 1930  
Db 354 GGCATCTCTGGGACCTGGAATCTATCACCCAGGAACTCTCAACCTCTTGCCCTGGAAC 295  
Qy 1931 agggccagataaggaacacactcgagccactttttgaagagcaatgtgaggaagagag 1990  
Db 294 AGGGCCAGATAAGGAACACCTCGGGCCACTCTTCTGAAGGCCAACCTGGAGGAAGGGAG 235  
Qy 1991 cagccagccgtttggagagaagatctcaagatccagactcattcctcctctgccc 2050  
Db 234 CAGCCAGCCATTTGGAGAGAGATCTCAAGATCCAGACTCATCTCTTCTCTGCCCC 175  
Qy 2051 agtgaattgtctctcctcagacttggggagctcctccttgaacccataagaccoca 2110  
Db 174 AGTAATTTGGTCTCTCCAGCTCTGGGGAGACTCTTCTTGAACCTTAATAAGACCCA 115  
Qy 2111 ctggagctcctctctcctcattccatccctcctctgcctctgctcttaattctgcagat 2170  
Db 114 CTGGAATCTCTCTCTCCATCCCTCTCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 55  
Qy 2171 gtcaaccacacactcactcgtgactcatataaataaagatatttcc 2224  
Db 54 GTCACCTCAAACTTACTCTGAGCTATTAATAATAACAGATTATTCTTC 1

RESULT 8  
W22160 636 bp mRNA EST 06-MAY-1996  
LOCUS W22160  
DEFINITION 63A6 Human retina cDNA Tsp509T-cleaved sublibrary Homo sapiens cDNA  
not directional, mRNA sequence.  
ACCESSION W22160  
VERSION W22160.1 GI:1299007  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 636)  
AUTHORS Macke, J., Smallwood, P. and Nathans, J.  
JOURNAL Adult Human Retina cDNA  
COMMENT Unpublished (1996)  
On Apr 14, 1993 this sequence version replaced gi:785820.  
Contact: Dr. Jeremy Nathans  
Dr. Jeremy Nathans, Dept. of Molecular Biology and Genetics  
Johns Hopkins School of Medicine  
725 North Wolfe Street, Baltimore, MD 21205  
Tel: 410 955 4678  
Fax: 410 614 0827  
Email: jeremy.nathans@jhu.edu  
Clones from this library are NOT available.  
PCR Primers  
FORWARD: CTTTGGACAGTTCAGCTCGGTAACT  
REVERSE: GAGTGGCTTATGATGTTCTTCTCCAGGTTAA  
Seq primer: GGGTAAAGCAAAAGATT.  
Location/Qualifiers  
1..636  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Human retina cDNA Tsp509T-cleaved sublibrary"  
/sex="mixed (males and females)"  
/tissue\_type="retina"  
/dev\_stage="adult"



REFERENCE 1 (bases 1 to 411)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
On Mar 16, 1998 this sequence version replaced gi:2961764.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)

FEATURES  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1..411  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2285306"  
/clone\_1lb="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: Kidney; Vector: p773D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(clonoids 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 95 a 83 c 131 g 102 t  
ORIGIN

Query Match 18.0%; Score 403; DB 49; Length 411;  
Best Local Similarity 98.8%; Pred. No. 1e-90;  
Matches 406; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1808 cccacctgtgaaacaaagcctcgagccttggagagggccacatgacacacacca 1867  
|||||  
411 CCCACCTCGTGAACACAAAGCCTCGGACAGCCTTGAGAGGCCACACTACACACACCA 352  
1868 gatggcatccttggagcctgaatctatccaccaggaatcacaactcccttggccctga 1927  
|||||  
351 GATGGCATCTTGGAGCAGCTGAATCTATCACCCAGAAATCTCAAACTCCCTTGGCCCTGA 292  
1928 accagggcagataaggaacagctcgagcacttttgaagggccaagtgtgaggaag 1987  
|||||  
291 ACCAGGGCCAGATAAGGAACAGCTCGGGCCACTCTTAAGGCCAATGTGAGGAAAG 232  
1988 gaggcagccagccggttggaggaagatcctaagatccagaactctatcttccctctg 2047  
|||||  
231 GAGCAGCCAGCGTTGGAGAGAGATCTCAAGATCCAAGCTCTCAATCTCTTCCTCTG 172  
2048 ccagatgaattgtctcccccagcttgggagccttcccttgaaccttaataagacc 2107  
171 CCCAGGAATTTGGTCTTCGCCAGCTCTGGGGAGACTCTTCCCTGGAACCTTAATAAGACC 112  
2108 ccactggaatctctctctcctcctcctctcctctcctctgccccttgctctaaattgctgcag 2167  
111 CCAGTGGAGATCTCTCTCTCCATCCCTCTGCGCTCCCTGCTTAATGTGTCGACG 52  
2168 attgtcaatccaaactctctgagctcaatataataaacagattta 2218  
51 ATTGTCACTCCAAACCTTACTCTGAGCTCATTAATAATAACAGATTTA 1

RESULT 11  
LOCUS AI989344/c  
DEFINITION wzz27e03.x1 NCI\_CGAP\_Brn53 Homo sapiens cDNA clone IMAGE:2559292 3',  
mRNA sequence.  
ACCESSION AI989344  
VERSION AI989344.1 GI:5836225  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 407)  
AUTHORS NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BRGAP), Tumor Gene Index  
JOURNAL Unpublished (1998)  
On Feb 18, 1999 this sequence version replaced gi:4296517.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life  
Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The  
I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)

FEATURES  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1..407  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2559292"  
/clone\_1lb="NCI\_CGAP\_Brn53"  
/tissue\_type="three pooled meningiomas"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: PCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies."

BASE COUNT 97 a 81 c 127 g 102 t  
ORIGIN

Query Match 17.6%; Score 394.2; DB 63; Length 407;  
Best Local Similarity 98.0%; Pred. No. 1.6e-88;  
Matches 399; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1818 gaacacaaagcctcgagccttgaagagggccacatgacacacacagatgcatcc 1877  
|||||  
407 GAACACAAAGCCTCGGATCAGCTTGAGAGAGGCCACACTACACACAGATGGCATCC 348  
1878 ttggagacctgaatctataccccaaggaatcctaactcccttggcccttaaacaggcca 1937  
|||||  
347 TTGGAGCTGAATCTATCACCCAGGAATCTCAAACTCCCTTGGCCCTTAACCAAGGCCCA 288  
1938 gataaagaacagctcgagcacttttgaagggccaatgtgaggaaggaaggaacagc 1997  
|||||  
287 GATAAGAACAGCTCGGGCCACTCTTGAAGGCCCAAGCTGAGGAAGGAAGGACAGCCAG 228  
1998 ccgattggagaagatcctaaggaatccagactcattccctctcctgcccagtgaa 2057  
111 CCATTGGGAGAGATCTCAAGATCCAAGATCAAGCTCTTCCCTGCGCCAGTGAAT 168  
227 CCATTGGGAGAGATCTCAAGATCCAAGATCAAGCTCTTCCCTGCGCCAGTGAAT 168  
2058 ttggtctctcccaagcttggggagactccttcttgaagccctataaagacccactggag 2117  
167 TTGCTCTCCAGCTCTGTGGGAGACTCTTCTTGAACCTTAATAAGACCCACTGAGAT 108  
2118 cctctctctcctcctctcctctcctctcctctcctctcctctcctctcctcctc 2177

QY	107	2178	47	RESULT 12	
Db	107	2178	47	AI196741/c	
	CTCTCTCTGCGATCCCTCTCTCTGCGCTGCTCTAATGCTGCGAGATGTCACATC	caaaccttactctgagctcattataataaataacagattatttcc	CAAACTTACTCTGAGCTCATTAATAATAATAAACAAGATTATTATTC	AI196741/c	
	LOCUS	LOCUS	LOCUS	LOCUS	
	DEFINITION	DEFINITION	DEFINITION	DEFINITION	
	ACCESSION	ACCESSION	ACCESSION	ACCESSION	
	VERSION	VERSION	VERSION	VERSION	
	KEYWORDS	KEYWORDS	KEYWORDS	KEYWORDS	
	SOURCE	SOURCE	SOURCE	SOURCE	
	ORGANISM	ORGANISM	ORGANISM	ORGANISM	
	REFERENCE	REFERENCE	REFERENCE	REFERENCE	
	AUTHORS	AUTHORS	AUTHORS	AUTHORS	
	TITLE	TITLE	TITLE	TITLE	
	JOURNAL	JOURNAL	JOURNAL	JOURNAL	
	COMMENT	COMMENT	COMMENT	COMMENT	
	FEATURES	FEATURES	FEATURES	FEATURES	
	source	source	source	source	
	16.7%	Score 372.8;	DB 60;	Length 384;	
	Best Local Similarity	98.2%;	Pred. NO. 3.7e-83;		
	Matches 377;	Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;
	1846	agaagccacatgacacacacacagatgacatctcttggagcttgaatcatalcaccaggaat	1905		
	384	AGAGGCCACACTACACACACACAGATGGATCTCTGGAGCTGAATCTATACACCGAAT	325		
	1906	ctcaaatccctcttggccttgacgaagggccagataaaggaaagctcgggcacttttt	1965		
	334	CTCAAACTCTCTTGTGGCCCTGGAACACAGGGCCAGATTAAGGAACAGCTCGGGCAGCTTCT	265		
	BASE COUNT	95 a	75 c	120 g	94 t
	ORIGIN				
	Insert Length: 496	Std Error: 0.00			
	Seq primer: -40UP from Gibco.				
	Location/Qualifiers				
	1. .384				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:2298051"				
	/clone_id="NCI_CGAP_Ki411"				
	/lab_host="DH10B"				
	/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_Ki43 was prepared, and ss circles were made in vitro. Following HAP hybridization, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneds 1323376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."				

QY	1966	gaagggccaatctgtggaggaagaaggaaacacacagccgcttgggagaagatcctaaagatcca	2025
Db	264	GAAGGCCAACGCTGGAGGAGAAAGGAGACACCCAGCCACTTTGGGAGAGAATCTCAAGATGCCA	205
QY	2026	gactctcatcttccttcctctcgtggccagatgaatttggctctctccagcgttggggagctcc	2085
Db	204	GACTCTCATCTTAACTGCTCTGCGCCAGAGTAATTTGGTCTCTCCAGCTCTGGGGAGCTCC	145
QY	2086	ttctcttgaacctataagaagaccacatgagatctctctctctccatccctctctctctc	2145
Db	144	TTCTCTTGAACCCCTAATAAGACCCCACTGAGATCTCTCTCTCCATCCCTCTCTCC	85
QY	2146	ctctgctctaaattgtgcgcagatgtcaccacccaactctactctgagctatnaataa	2205
Db	84	CTCTGCTTAATTTGCTGTGCCAGAGATTGTACTTCACCAACTTACTGTGAGCTATTAATAA	25
QY	2206	ataacagattatatttccagctt 2229	
Db	24	ATAACAGATTATTTTCCAGCTT 1	
RESULT	13		
LOCUS	A1272081/c		
DEFINITION	A1272081 385 bp mRNA EST 17-MAR-1999		
ACCESSION	q178f10.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1865611 3',		
VERSION	mRNA sequence.		
KEYWORDS	A1272081 A1272081.1 GI:3891248		
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
TITLE	Eutheria; Primates; Catarrhini; Homiidae; Homidae; Homo.		
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
	Tumor Gene Index		
	Unpublished (1997)		
	On Jan 14, 1998 this sequence version replaced gi:1197673.		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.		
	Emmert-Buck, M.D., Ph.D.		
	CDNA Library Preparation: M. Bento Soares, Ph.D.		
	CDNA Library Arrayed by: Greg Lennon, Ph.D.		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNLB at:		
	www-bio.lnl.gov/bbrp/Image/Image.html		
FEATURES			
SOURCE			
	Insert Length: 568 Std Error: 0.00		
	Seq primer: -40UP from GlbCO		
	High quality sequence stop: 379.		
	location/Qualifiers		
	1. 385		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:1865611"		
	/clone_1lb="NCI CGAP_Kid3"		
	/lab_host="DH10B"		
	/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with		
	a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st		
	strand cDNA was primed with a Not I - oligo(dT) primer,		
	double-stranded cDNA was ligated to Eco RI adaptors		
	(Pharmacia), digested with Not I and cloned into the Not		
	I and Eco RI sites of the modified pT73 vector. mRNA		
	source: 2 pooled kidneys. Library went through one round		
	of normalization. Library constructed by Bento Soares and		
	M. Fatima Bonaldo.		
BASE COUNT	94 a 72 c 118 g 101 t		
ORIGIN			





Seq primer: -40UP from Gibco  
High quality sequence stop: 338.  
Location/Qualifiers  
source  
1. 386

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2143700"  
/clone.lib="NCI CGAP Pan1"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: pancreas; Vector: pCMV-SPORT6; site\_1: SalI;  
site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.72 kb. Life Technologies catalog #:  
11548-013"

BASE COUNT 96 a 71 c 116 g 102 t 1 others  
ORIGIN

Query Match 16.5%; Score 369.2; DB 46; Length 386;  
Best Local Similarity 97.7%; Pred. No. 2.9e-82;  
Matches 374; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1854 catcacacacacacagatgcatccttggacatcctatcaccacagaaatcctaact 1913  
|||  
Db 386 CACTACACACACAGATGGCATCTGGGACCTGATCTATCACCGAGATCTCAACT 327  
QY 1914 ccccttggccctgaaccagagccagataagaagctcgagccactttttgaagcca 1973  
|||  
Db 326 CCCCTTGGCCCTGAACAGGGCCAGATAAGGACAGCTCGGGCCACTTCTGAAGGCCA 267  
QY 1974 atgtggaggaaggaagcagccagccgtttggaggaagatcctcaagatccagactctca 2033  
|||  
Db 266 ACGTGAAGGAAGGAGGAGCAGCCACCTTTGGAGAAGATCTCAAGATCCAGACTCTCA 207  
QY 2034 ttccttccctcctggccagtgaaattgtctctcctccagcttggggagactcctccttga 2093  
|||  
Db 206 TTCCTTCCCTGGCCAGTGAATTGGTCTCTCCAGCTTGGGGAGCTCTCTTGA 147  
QY 2094 accctaataagaaccacacagatcctctctctccacccctcctcctcctcgtc 2153  
|||  
Db 146 ACCCTAATAAGACCCCACTGGAGTCTCTCTCCATCCCTCTCTGCCCCCTGCTT 87  
QY 2154 taattgtgcagagatgtcactccaaccttactctgagctcattataataataaag 2213  
|||  
Db 86 TAATTGCTGCCAGGATGTCTCAACCACTTACTGTGAGCTCATTAATAATAAACAG 27  
QY 2214 attatttccagcttaaaaaa 2236  
|||  
Db 26 ATTATTTTCCAGCTTAAAAAA 4

Search completed: May 19, 2000, 22:33:48  
Job time: 9649 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 00:27:48 ; Search time 158.18 Seconds  
(without alignments)  
2315.815 Million cell updates/sec

Title: US-09-502-945-2  
Perfect score: 2885  
Sequence: 1 ggaattccctctgtcgaagt.....aaacaaaagttaaaattt 2885

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 226296 segs, 63486255 residues

Total number of hits satisfying chosen parameters: 452592

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued Patents, NA:\*

- 1: /cgn2-6/ptodata/2/ina/5A.COMB.seq:\*
- 2: /cgn2-6/ptodata/2/ina/5B.COMB.seq:\*
- 3: /cgn2-6/ptodata/2/ina/5C.COMB.seq:\*
- 4: /cgn2-6/ptodata/2/ina/5D.COMB.seq:\*
- 5: /cgn2-6/ptodata/2/ina/6.COMB.seq:\*
- 6: /cgn2-6/ptodata/2/ina/PCrus.COMB.seq:\*
- 7: /cgn2-6/ptodata/2/ina/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	84.6	2.9	3489	4	US-08-728-323A-1
2	84.6	2.9	32207	3	US-08-770-379-20
3	79	2.7	5452	4	US-09-130-114-1
4	79	2.7	10596	1	US-07-884-811-15
5	79	2.7	10596	1	US-07-885-971-15
6	79	2.7	10596	1	US-08-087-783A-15
7	79	2.7	10596	1	US-08-194-088B-15
8	79	2.7	10596	3	US-08-194-087-15
9	79	2.7	10596	6	PCT-US93-04648-15
10	78.8	2.7	7218	4	US-08-232-463-14
11	72.6	2.5	1931	4	US-09-130-114-2
12	65	2.3	234	2	US-08-469-802B-3
13	65	2.3	234	2	US-08-267-803B-3
14	62	2.1	2338	2	US-08-425-069-1
15	62	2.1	2338	4	US-08-317-844B-1
16	60.4	2.1	2793	2	US-08-209-747-1
17	60.4	2.1	2793	2	US-08-458-298-1
18	60.2	2.1	543	7	5273901-6
19	59	2.0	1505	1	US-07-915-246-1
20	58.8	2.0	171	2	US-08-469-802B-5
21	58.8	2.0	171	3	US-08-267-803B-5
22	58.8	2.0	195	2	US-08-469-802B-2
23	58.8	2.0	195	3	US-08-267-803B-2
24	57	2.0	168	2	US-08-469-802B-4
25	57	2.0	168	3	US-08-267-803B-4
26	56.8	2.0	1995	2	US-08-425-069-3
27	56.8	2.0	1995	4	US-08-317-844B-3

28	56	1.9	9551	1	US-08-056-200-93	Sequence 93, Appl
29	56	1.9	9551	4	US-08-800-644-93	Sequence 93, Appl
30	55.4	1.9	2214	5	US-08-864-038A-1	Sequence 1, Appl
31	55.4	1.9	3331	5	US-08-864-038A-2	Sequence 2, Appl
32	55.4	1.9	3331	5	US-08-864-038A-4	Sequence 4, Appl
33	55	1.9	1052	2	US-08-466-603-1	Sequence 1, Appl
34	55	1.9	1052	2	US-08-314-503A-1	Sequence 1, Appl
35	55	1.9	1052	2	US-08-468-066-1	Sequence 1, Appl
36	55	1.9	1052	3	US-08-466-717-1	Sequence 1, Appl
37	55	1.9	1052	6	PCT-US95-12414-1	Sequence 1, Appl
38	54	1.9	336	4	US-07-814-220-4	Sequence 4, Appl
39	54	1.9	336	4	US-07-812-421-4	Sequence 4, Appl
40	54	1.9	533	7	5482709-5	Patent No. 5482709
41	53.6	1.9	12001	2	US-08-458-568A-11	Sequence 11, Appl
42	52.8	1.8	2818	1	US-08-366-276-1	Sequence 1, Appl
43	52.6	1.8	15378	5	US-08-785-420-1	Sequence 1, Appl
44	52.4	1.8	336	4	US-07-814-220-3	Sequence 3, Appl
45	52.4	1.8	336	4	US-07-812-421-3	Sequence 3, Appl

## ALIGNMENTS

RESULT 1  
US-08-728-323A-1  
; Sequence 1, Application US/08728323A  
; Patent No. 5948676  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Yuan  
; APPLICANT: Bohenzky, Roy A.  
; APPLICANT: Russo, James J.  
; APPLICANT: Eelman, Isidore S.  
; APPLICANT: Moore, Patrick S.  
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA  
; TITLE OF INVENTION: Encoding Same And Uses Thereof  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/728, 323A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MS/CS/SKS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-278-0400  
; TELEFAX: 212-391-0525  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3489 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3489  
; US-08-728-323A-1

	Query Match	2.9%	Score 84.6:	DB 4:	Length 3489:
	Best Local Similarity	48.8%:	Pred. No. 7.4e-11:		
	Matches 293:	Conservative	0:	Mismatches 299:	Indels 9:
				Gaps	2:
QY	ctcgacgatgtctgtctgtcgtgagcaggcccccgcagcaagaccctcatctgttgcga	829			
Db	1603 CAGCAGCAGGAGACCCACAGCAGCAGGAGGCCACAGCAGCAGGAGCCACAGCAGGGAGGCCA	1662			
QY	830 ctccagggcagttccccactagtgaacggttgaaagltgttgccaagaatcgcggaagta	889			
Db	1663 CAGCAGCGGGAGCCCCCAGCAGCGGGAGGCCACAGCAGCGGGAGGCCACAGCAGCGGGAGGCCA	1722			
QY	890 ggcgaagctcccgccgcatctcgcgcccttgagccgactaatgttcctcacgctgcgcgaag---	947			
Db	1733 CAGCAGCGGGAGCCACAGCAGCGGGAGGCCACAGCAGCGGGAGGCCACAGCAGCGGGAGGCCA	1782			
QY	947 agtcccaagggcccttcagcagcttgttcatltyacaacaacagcaccacagcttcttgagaag	1006			
Db	1783 CAGCAGCAGGATGAGACACAGCAGCAGGATGACAGCAGCAGGATGAGACACAGCAGGATGAG	1842			
QY	1007 cagaagcagcagcagctacagcttggcaagatcctcaaccaagaacagggagctgtgccag	1066			
Db	1843 CAGCAGCAGGATGAGACAGCAGCAGGATGAGCAGCAGGATGAGACAGCAGCAGGATGAG	1902			
QY	1067 cagcccaaccacccaccccttgagagacagagagagcttgcagagcagcagagagctctg	1126			
Db	1903 CAGCAGCAGGATGAGACAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGGATGAG	1962			
QY	1127 ctgggggagggagggcccttgaccatgcccgggaggggctccacagagatgagagacacag	1186			
Db	1963 CAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGGATGAG	2016			
QY	1187 gaagacctggagagagagacgaagaaagagattgggagagagagatgtacccag	1246			
Db	2017 GATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCAG	2076			
QY	1247 gttaaagcagcagagagggagagatgtgtcttgaggaggggccgacttggagagacctgt	1306			
Db	2077 GATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAG	2136			
QY	1307 gctgatacacaanaactgttctcagatgcccacacgctgcaaccttgcaggtgtaccaa	1366			
Db	2137 GAGCAGCAGGATGAGACAGCAGCAGGATGAGCAGCAGCAGGATGAGACAGCAGCAGCAG	2196			
QY	1367 g 1367				
Db	2197 G 2197				
<b>RESULT 2</b>					
US-08-770-379-20/c					
Sequence 20, Application US/0870379					
Patent No. 5849564					
<b>GENERAL INFORMATION:</b>					
APPLICANT: Chang, Yuan					
APPLICANT: Bohenzky, Roy A.					
APPLICANT: Russo, James J.					
APPLICANT: Edelman, Isidore S.					
APPLICANT: Moore, Patrick S.					
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED					
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF					
NUMBER OF SEQUENCES: 20					
CORRESPONDENCE ADDRESSES:					
ADDRESS: Cooper & Dunham LLP					
STREET: 1185 Avenue of the Americas					
CITY: New York					
STATE: New York					
COUNTRY: U.S.A.					
ZIP: 10036					
COMPUTER READABLE FORM:					
MEDIUM TYPE: Floppy disk					
COMPUTER: IBM PC compatible					
OPERATING SYSTEM: PC-DOS/MS-DOS					

```

1 SOFTWARE: PatentIn Release #1.0, Version #1.30
2
3 CURRENT APPLICATION DATA:
4 APPLICATION NUMBER: US/08/770,379
5
6 FILING DATE:
7 CLASSIFICATION: 435
8
9 ATTORNEY/AGENT INFORMATION:
10 NAME: White, John P.
11
12 REGISTRATION NUMBER: 28,678
13 REFERENCE/DOCKET NUMBER: 52342
14
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE: (212) 278-0400
17 TELEFAX: (212) 391-0525
18
19 INFORMATION FOR SEQ ID NO: 20:
20
21 SEQUENCE CHARACTERISTICS:
22
23 LENGTH: 32207 base pairs
24 TYPE: nucleic acid
25 STRANDEDNESS: double
26
27 TOPOLOGY: linear
28
29 MOLECULE TYPE: DNA (genomic)
30
31 US-08-770-379-20

```

Query Match	Similarity	2.9%	Score 84.6	DB 3	Length 32207
Best Local	Similarity	48.8%	Pred. No. 1.8e-10		
Matches	293	Conservative	0	Mismatches	299
				Indels	9
				Gaps	2
QY	770	ctcagcagcatgtgtctgtcgtgagcagagcccgccagcagcagagaccctcattgtctgtccca	829		
Db	20394	CAGCAGCAGGAGGAGCCACAGCAGCAGGAGGAGCCACAGCAGCAGGAGCAGCAGCAGGAGGAGC	20335		
QY	830	ctccacagggcaggtctcccccattgacccgggtgaagctgtgtgccaccagcatgtggagcgtga	889		
Db	20334	CAGCAGCGGGAGCCCCACAGCAGCGGAGCCACAGCAGCGGGAGCCACAGCAGCGGGAGC	20275		
QY	890	ggcaagctcccccggcgcatcgtgcgccttgagcgcgactcaagctctccaccgctgcgcgac---	947		
Db	20274	CAGCAGCGGGAGGAGCCACAGCAGCGGGAGCCACAGCAGCGGGAGCCACAGCAGCGGGAGC	20215		
QY	947	agtcgcccaagcctctgcagcagctgtgtcaltgcacaacacagcaccagcagtcctctggagaag	1006		
Db	20214	CAGCAGCAGGAGTGAACAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAG	20155		
QY	1007	cagaagcagcagcagctacagctgtggcgaagatcctccaccagaacagggagctgtgccag	1066		
Db	20154	CAGCAGCAGGAGTGAACAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAG	20095		
QY	1067	cagccaccacccacccctctgagagacagagagagcgtcagcagcagcagcagcgtcttg	1126		
Db	20094	CAGCAGCAGGAGTGAACAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAG	20035		
QY	1127	ctggggagagagagcccttgaccatctgcgccggagagggctccacagagagtgagagcacag	1186		
Db	20034	CAGCAGCAGGAGTGAACAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAG	1981		
QY	1187	gaagaccctggagagagagacgagagagaagagatgtggagagagagagagatttcacccag	1246		
Db	19980	GATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGCAG	19921		
QY	1247	gttaagagcagagagagggagagagtggtgtctggagagaggggcccagcttgagagagccttgt	1306		
Db	19920	GATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAG	19861		
QY	1307	gctgatatcaaaaaactgttctcagatgtcccaacgcgtctgcaactcttgcaagtgtaacca	1366		
Db	19860	GAGCAGCAGGAGTGAACAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAG	19801		
QY	1367	g 1367			
Db	19800	G 19800			

; Sequence 1, Application US/09130114  
; Patent No. 5976807  
; GENERAL INFORMATION:  
; APPLICANT: Horlick, Robert A.  
; APPLICANT: Damaj, Bassam B.  
; APPLICANT: Robbins, Alan K.  
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes  
; TITLE OF INVENTION: From Multiple Transfected Epilosomes  
; FILE REFERENCE: 0867/1D903US1  
; CURRENT APPLICATION NUMBER: US/09/130,114  
; CURRENT FILING DATE: 1998-08-06  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 5452  
; TYPE: DNA  
; ORGANISM: YEBNA  
US-09-130-114-1

Query Match 2.7%; Score 79; DB 4; Length 5452;  
Best Local Similarity 51.3%; Pred. No. 1.8e-09;  
Matches 184; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 943 gcaagagtcctccagccctgcaagcgtgtcatgtcaacaacacacacagtccttga 1002  
DB 2016 GCAGAGAGGGGCGAGAGGGGCGAGAGAGGGGCGAGAGAGGGGCGAGCA 1957  
QY 1003 gaagcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1062  
DB 1956 GGGGCGAG 1897  
QY 1063 cagcagcagccaccaccaccctgtgagagagagagagagagagagagag 1122  
DB 1896 GGAGCGAG 1837  
QY 1123 ctgtcgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1182  
DB 1836 GCAGCA 1777  
QY 1183 acagaagaacccctggaagagagagagagagagagagagagagagagagat 1242  
DB 1776 GGAGGGGCGAG 1717  
QY 1243 ccaggttaagagcagagagagagagagagagagagagagagagagagagc 1301  
DB 1716 GGAGCGAGC 1658

RESULT 4  
US-07-884-811-15  
; Sequence 15, Application US/07884811  
; Patent No. 5316921  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/884,811  
; FILING DATE: 19920518  
; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Ginger R.  
; REGISTRATION NUMBER: 33,055  
; REFERENCE/DOCKET NUMBER: 755.1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-3216  
; TELEFAX: 415/952-9881  
; TELEEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10596 bases  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-07-884-811-15

Query Match 2.7%; Score 79; DB 1; Length 10596;  
Best Local Similarity 51.3%; Pred. No. 2.4e-09;  
Matches 184; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 943 gcaagagtcctccagccctgcaagcgtgtcatgtcaacaacacacacagtccttga 1002  
DB 2369 GCAGAGAGGGGCGAGAGAGGGGCGAGAGAGGGGCGAGAGAGGGGCGAGCA 2428  
QY 1003 gaagcagaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1062  
DB 2429 GGGGCGAG 2488  
QY 1063 cagcagcagccaccaccaccctgtgagagagagagagagagagagagag 1122  
DB 2489 GGAGCGAG 2548  
QY 1123 ctgtcgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtg 1182  
DB 2549 GCAGCA 2608  
QY 1183 acagaagaacccctggaagagagagagagagagagagagagagagagagat 1242  
DB 2609 GGAGGGGCGAG 2668  
QY 1243 ccaggttaagagcagagagagagagagagagagagagagagagagagagc 1301  
DB 2669 GGAGCGAGC 2727

RESULT 5  
US-07-885-971-15  
; Sequence 15, Application US/07885971  
; Patent No. 5328837  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/885,971  
; FILING DATE: 19920518  
; CLASSIFICATION: 530





```

: FILING DATE: 19930517
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/884811
: FILING DATE: 18-MAY-92
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/885971
: FILING DATE: 18-MAY-92
: ATTORNEY/AGENT INFORMATION:
: NAME: Dreger, Ginger R.
: REGISTRATION NUMBER: 33,055
: REFERENCE/DOCKET NUMBER: 755,779P1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-3216
: TELEFAX: 415/952-9881
: TELE: 910/371-7168
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10596 bases
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: PCT-US93-04648-15

```

```

Query Match          2.7%; Score 79; DB 6; Length 10596;
Best Local Similarity 51.3%; Pred. No. 2.4e-09;
Matches 184; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

```

```

Qy 943 gcagagtcacagccctcagcagctgctatgcaacaacagcagctcctcga 1002
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2369 gcagagtcacagccctcagcagctgctatgcaacaacagcagctcctcga 2428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1003 gaagcagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1062
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2429 gggcagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2488
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1063 cagcagcccccaccccccctcgaagacagcagcagcagcagcagcagcagc 1122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2489 gcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2548
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1123 ctgctgaggagagagcctcgaacatgccccggagagcctccacagagctgagagc 1182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2549 gcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2608
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1183 acagagacactgagagagagcagcagcagcagcagcagcagcagcagcagc 1242
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2609 gcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2668
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1243 ccaggttaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2669 gcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2727
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 10
US-08-232-463-14/c
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

```

: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232.463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935.313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELE: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: pTZgpt-Fls
: US-08-232-463-14

```

```

Query Match          2.7%; Score 78.8; DB 1; Length 7218;
Best Local Similarity 3.6%; Pred. No. 2.3e-09;
Matches 14; Conservative 241; Mismatches 133; Indels 0; Gaps 0;

```

```

Qy 963 agcagctgctatgcaacaacagcagcagcagcagcagcagcagcagcagcagc 1022
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1428 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1369
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 1023 tacagctggcagagctcctccacagacagcagcagcagcagcagcagcagcagc 1082
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1368 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1309
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 1083 ctgagagacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1142
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1308 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1249
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 1143 tgacatccccccggagcctccacagagctgagcagcagcagcagcagcagcagc 1202
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1248 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1189
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 1203 agcagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1262
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1198 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1129
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 1263 gcagagctgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1322
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1128 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1069
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Qy 1323 tgtctcagatgccacacgcagcagcagcagcagcagcagcagcagcagcagc 1350
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 1068 RRATCGCAAGCTCCCTCGACCTGCAGCC 1041
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

```

RESULT 11
US-09-130-114-2/c
: Sequence 2, Application US/09130114
: Patent No. 5976807
: GENERAL INFORMATION:
: APPLICANT: HORLICK, Robert A.
: APPLICANT: DAMEJ, Bassam B.
: APPLICANT: ROBBINS, Alan K.
: TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes

```



```

? TITLE OF INVENTION: From Multiple Transfected Episomes
?
? FILE REFERENCE: 0867/1D9030U1
?
? CURRENT APPLICATION NUMBER: US/09/130,114
?
? CURRENT FILING DATE: 1998-08-06
?
? NUMBER OF SEQ ID NOS: 36
?
? SOFTWARE: FastSeq for Windows Version 3.0
?
? SEQ ID NO 2
?
? LENGTH: 1931
?
? TYPE: DNA
?
? ORGANISM: EBNA
?
? US-09-130-114-2

```

Query Match	2.5%	Score 72.6;	DB 4;	Length 1931;
Best Local Similarity	48.6%;	Pred. No. 3.8e-08;		
Matches 198; Conservative	0;	Mismatches 209;	Indels 0;	Gaps 0;

[illegible]

RESULT 12  
 US-08-469-802B-3  
 Sequence 3, Application US/08469802B  
 Patent No. 5741645  
 GENERAL INFORMATION:  
 APPLICANT: Orr, Harry T.  
 APPLICANT: Rannu, Laura P.W.  
 APPLICANT: Chung, Ming-Yi  
 APPLICANT: Zoghbi, Huda Y.  
 TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia  
 Patent No. 5741645  
 TITLE OF INVENTION: Type 1 and Method for Diagnosis  
 NUMBER OF SEQUENCES: 47  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Muelting, Raasch, Gebhardt & Schnappach, P.A.  
 STREET: 119 No. 5741645th Fourth Street, Suite 203  
 CITY: Minneapolis  
 STATE: MN  
 COUNTRY: USA  
 ZIP: 55401  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/469, 802B

```

1      FILING DATE: 06-JUN-1995
2      CLASSIFICATION: 435
3      ATTORNEY/AGENT INFORMATION:
4      NAME: Meeting, Ann M.
5      REGISTRATION NUMBER: 33,977
6      REFERENCE/DOCKET NUMBER: 110.00030101
7      TELECOMMUNICATION INFORMATION:
8      TELEPHONE: 612-305-1217
9      TELEFAX: 612-305-1225
10     INFORMATION FOR SEQ ID NO: 3:
11     SEQUENCE CHARACTERISTICS:
12     LENGTH: 234 base pairs
13     TYPE: nucleic acid
14     STRANDEDNESS: single
15     TOPOLOGY: linear
16     MOLECULE TYPE: DNA
17     OS-08-469-802b-3

```

Query Match	2.38;	Score 65;	DB 2;	Length 234;
Best Local Similarity	55.68;	Pred. No. 1e-06;		
Matches 125;	Conservative 0;	Mismatches 100;	Indels 0;	Gaps 0;

[illegible]

RESULT 13  
 US-08-267-803B-3  
 Sequence 3, Application US/08267803B  
 Patent No. 5834183  
 GENERAL INFORMATION:  
 APPLICANT: Orr, Harry T.  
 APPLICANT: Rannu, Laura P.W.  
 APPLICANT: Chung, Ming-Yi  
 APPLICANT: Zoghbi, Huda Y.  
 TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia  
 Patent No. 5834183  
 TITLE OF INVENTION: Type 1 and Method for Diagnosis  
 NUMBER OF SEQUENCES: 85  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Muelten, Raasch, Gebhardt & Schwappach, P.A.  
 STREET: P.O. Box 581415  
 CITY: Minneapolis  
 STATE: MN  
 COUNTRY: USA  
 ZIP: 55458-1415  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/267, 803B  
 FILING DATE: 28-JUN-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: McCormack, Myra H.  
 REGISTRATION NUMBER: 36,602  
 REFERENCE/DOCKET NUMBER: 110.00030120



```

QY 1235 gattgcatccaggtttaaagacagcagagagagagcagagagtgtgtccgagagagagagccccaattg 1294
Db 944 GTGCAAGGTGCAAGCAGCAGCTGTGAGAGGTGCCGAGCAAGAGGACTAGGTGACAAAGGTGCTG 1003

QY 1295 gaagagag 1300
Db 1004 GACAAAG 1009

Search completed: May 20, 2000, 00:30:22
Job time: 11605 sec

```

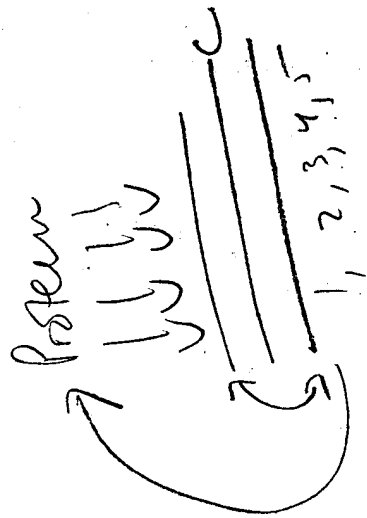
```
Search completed: May 20, 2000, 00:30:22
Job time: 11605 sec
```

```

QY 1235 gattgcattccagggttaagagacagagagagggcgagagatggtgctgagagagggcccgacttg 1294
Db 944 GTGCAGGTGCAGCAGCAgCTTGAGAGGTGCCGCGCAGAGGAGGACTTAAGGTGAGACAAGGTGCTG 1003
QY 1295 gaaggag 1300
Db 1004 GACCAAG 1009

```

Search completed: May 20, 2000, 00:30:22  
Job time: 11605 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 19, 2000, 19:52:59 ; Search time 3256.04 Seconds  
(without alignments)  
1931.980 Million cell updates/sec

Title: US-09-502-945-1  
Perfect score: 1552  
Sequence: 1 ctctcgatgcacgcgagaa.....aaatgaacttttaagaaa 1552

Scoring table: IDENTITY\_NDC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 segs, 202611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result

\*  
Query

No.	Score	Match	Length	DB	ID	Description
C 1	465.2	30.0	490	51	A1735499	A1735499 a112006.x.s1
C 2	412	26.9	479	64	A10028371	A10028371 wv88509.x.s1
C 3	371.4	23.5	410	46	A1445451	A1445451 t_j34b11.x.s1
C 4	369	23.8	391	42	A1092201	A1092201 qa60049.s.s1
C 5	349.4	22.5	365	44	A10266387	A10266387 qm45008.x.s1
C 6	345.4	22.3	360	44	A12868955	A12868955 q181c08.x.s1
C 7	335.8	21.6	398	39	AA886109	AA886109 y144f05.s.s1
C 8	328.4	21.2	433	42	A1140814	A1140814 ool15c05.x.s1
C 9	327.2	21.1	434	41	A11088590	A11088590 ox34e03.s.s1
C 10	326.4	21.0	424	39	AA829894	AA829894 oes1e12.s.s1
C 11	326.4	21.0	430	39	AA879456	AA879456 o_j919f03.s.s1
C 12	326.4	21.0	439	27	AA029201	AA029201 zk12f08.s.s1
C 13	326.4	21.0	455	28	AA102109	AA102109 zk87g11.x.s1
C 14	326.4	21.0	462	43	A1168222	A1168222 ok30c01.x.s1
C 15	326.4	21.0	472	25	NS1485	NS1485 yz04e06.s1
C 16	326.4	21.0	482	42	A1148323	A1148323 qp31d07.x.s1
C 17	326.4	21.0	487	39	AA843811	AA843811 ak09c08.s.s1
C 18	326.4	21.0	489	44	A10298467	A10298467 qm90c10.x.s1
C 19	326.4	21.0	567	42	A1089322	A1089322 qp16d08.x.s1
C 20	326.4	21.0	596	26	W72147	W72147 zd70f08.s1
C 21	325.4	21.0	587	74	AA1923663	AA1923663 km29p02.x.s1
C 22	324.8	20.9	418	60	A10808317	A10808317 wf54d08.x.s1
C 23	324.8	20.9	442	41	A1033069	A1033069 ow93f02.s.s1
C 24	323.8	20.9	499	25	NS1277	NS1277 yz14d07.s1
C 25	323	20.8	482	74	AA1951519	AA1951519 km85b07.x.s1
C 26	321.8	20.7	556	29	AA1616455	AA1616455 zot73a06.s.s1
C 27	321.6	20.7	437	38	AA7442495	AA7442490 nv51d02.s.s1
C 28	320.6	20.7	572	63	A1970343	A1970343 wg91a04.x.s1
C 29	317.4	20.5	444	38	R71679	R71679 y185e08.s1
C 30	317.4	20.5	444	38	AA8182519	AA8182519 a1179p03.s1
C 31	315.4	20.3	422	25	NS4528	NS4528 yz91e06.s1
C 32	315.4	20.3	436	24	H99906	H99906 yx32h10.s1
C 33	313.4	20.2	472	28	AA099913	AA099913 zk87g11.x.s1
C 34	311.4	20.1	451	64	AA0021346	AA0021346 d12f1c04.y.s1
C 35	308.2	19.9	345	39	AA883684	AA883684 a158a05.s.s1
C 36	304.4	19.6	377	44	A1267979	A1267979 go44f190.x.s1
C 37	304.4	19.6	391	28	AA083859	AA083859 z16d06.s.s1
C 38	286.6	19.1	407	22	R36854	R36854 yf52c07.s1
C 39	292.8	18.9	451	28	AA089195	AA089195 mc05h11.r.s1
C 40	292.2	18.8	432	22	R39448	R39448 hc99s03.s1
C 41	288.2	18.6	372	32	AA34369	AA34369 est52776
C 42	287.2	18.5	468	24	H98684	H98684 yx17g01.s1
C 43	284.4	18.3	362	40	AA910762	AA910762 o125n06.s.s1
C 44	284.4	18.0	465	21	R07471	R07471 y937a06.s1
C 45	278.4	17.9	295	49	F25334	F25334 hspD12318 HSPD12318

## ALIGNMENTS

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40UP from Gibco  
High quality sequence stop: 344.

FEATURES  
Source  
Location/Qualifiers  
1. .490  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2354891"  
/clone\_lib="Barstead aorta HPLRB6"  
/sex="male"  
/dev\_stage="adult, age 64"  
/lab\_host="DH10B (phage resistant)"  
/note="Organ: aorta; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGAGCGCGCCGCTTTTCTTTTCTTTTCTTTTCTTTT  
3'], double-stranded cDNA was ligated to Eco RI adaptors  
[5', AATTGGATGGAAC 3' and 5', GTTGAATCGG 3'], digested  
with Not I and cloned into the Not I and Eco RI sites of  
the modified pT73 vector. Library constructed by Bob  
Barstead."

BASE COUNT 111 a 130 c 119 g 130 t  
ORIGIN

Query Match 30.0%; Score 465.2; DB 51; Length 490;  
Best Local Similarity 98.2%; Pred. No. 5.5e-110;  
Matches 481; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

07 1060 gggaagagactctgagacagatgaagcaaaagcgctaaagcgctgagtaagcacagcgcc 1119  
Db 490 gggagagatcattgaaacacatgaagcaaaagggctaaagcgacgttgatnagcacagcaggcc 431  
09 1120 acagccagcagctgctgctcctcctcagcaagcagaacacagctctctctgagagcgag 1179  
Db 430 ACAGCCAGCAGCAGCTGTGTCAGCTCTTCAGCAAGCAAGCAACACACTTCTCTGTGAGAGCGAG 371  
09 1180 agcctgtcggaaagagtgtagccgctctgcggaccagttaccagcagtcacacatctgat 1239

RESULT	1
LOCUS	A1735499/c
DEFINITION	A1735499 490 bp mRNA EST 14-June-1999
ACCESSION	U182806.x1 Barstead aorta HPLRB6 Homo sapiens cDNA clone
VERSION	IMAGE:2354491.3 similar to tr:060527 060527 ANTIGEN NT-CO-8 ;,
KEYWORDS	mRNA sequence.
SOURCE	A1735499
ORGANISM	A1735499.1 GI:5057023
REFERENCE	EST.
AUTHORS	human.
TITLE	Homo sapiens
JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
COMMENT	Eutheria; Primates; Catarrhini; Homnidae; Homo.
	1 (bases 1 to 490)
	Hillier,L., Allen,M., Bowles,J., Dubouque,T., Geisels,G., Jost,S.,
	Kritzman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marrs,M.,
	Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
	Theising,B., White,Y., Wyllie,T., Waterston,R. and Wilson,R.
	Mashd-NCI human EST Project
	Unpublished (1997)
	On May 18, 1998 this sequence version replaced gi:3137107.

QY	1300	tggtttaaacttaataatagccacaacgacacacacccctccagggtgaaacccgctcagc	1359
Db	250	tggtttaaccttaataatagccacaacgacacacacccctccagggtgaaacccgctcagc	191
QY	1360	ctgcagtgaggctggttcctcaatcaacgctgggcgtgtccccgacgcaatcgagcttgag	1419
Db	190	ctgcagtgaggctggttcctcaatcaacgctgggcgtgtccccgacgcaatcgagcttgag	131
QY	1420	ctggaggttcgactcagcttgagcaga-ctccctgggtatgttttaagaataagctctgag	1478
Db	130	ctggaggttcgactcagcttgagcagaacgtccctgggtatgttttaagaataagctctgag	71
QY	1479	ttaatggtttaaatctgctcatctglatgctaggtatatacatatgatttccaataatga	1538
Db	70	ttatgtgttttaaatctgctcatctglatgctaggtatatacatatgatttccaataatga	11
QY	1539	actttttaa	1548
Db	10	actttttaa	1

RESULT 2  
AM028371/c 479 bp mRNA EST 15-SEP-1999  
LOCUS w988b09.x1 NCI\_CGAP\_Gas4 Homo sapiens cDNA clone IMAGE:2541209 3'  
DEFINITION similar to TR:060527 060527 ANTIGEN NY-CO-8 ;, mRNA sequence.  
ACCESSION AM028371  
VERSION AM028371.1 GI:5887127  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 479)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jun 22, 1998 this sequence version replaced gi:3247399.  
Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILM at:  
[www-bio.illn1.gov/bbrp/image/image.html](http://www-bio.illn1.gov/bbrp/image/image.html)

FEATURES  
source  
Seq primer: -40UP from Gibco  
High quality sequence stop: 361.  
Location/Qualifiers  
1..479  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2541209"  
/clone\_lib="NCI\_CGAP\_Gas4"  
/tissue\_type="poorly differentiated adenocarcinoma with  
signed ring cell features"  
/lab\_host="DH10B"  
/note="Organ: stomach; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.69 kb. Life Technologies catalog #:  
11549-011"

BASE COUNT 113 a 121 c 118 g 127 t  
ORIGIN

Query Match 26.5%; Score 412; DB 64; Length 479;  
Best Local Similarity 95.0%; Pred. No. 3.3e-96;  
Matches 458; Conservative 0; Mismatches 20; Indels 4; Gaps 3;

QY 1069 catgaagatgaagcaagctgaagcagctgataagcaagcagcagccagccag 1128  
|||||  
DB 479 CATGAGAGGATGAACCAAGCTTAGCAGCTTGAAATTA--GCCCAAGAGGCCACAGCCAGC 422  
|||||  
QY 1129 cagctgctgagctcctcctcagcagcagcctctcctctgagagcagcagcctgctg 1188  
|||||  
DB 421 AAGCGGTGCAAGCTCTCTGACGACGAAACCAAGCTTCTCTCTGAGAGGACAGAGCCT-TCG 363  
|||||  
QY 1189 gaagaggtggaacggctgctggaacccagttaccagcagtcgacaactctgattctgaact 1248  
|||||  
DB 362 GAAGAGGTGAGACCGGCTCGCGACCCAGTTACCCAGCATCTGATGCTGACCT 303  
|||||  
QY 1249 ggatggaaacagtgaaataatgaatatacaaaagatataatctatctcgtgttga 1308  
|||||  
DB 302 GGATGGAACAGAGTGAATTAATGATTTACAAAGAGATTTTACATCTCGGTTTGA 243  
|||||  
QY 1309 ctaataatgacacagcagcagcctcccaagggtgacacagcctcagcctgagtg 1368  
|||||  
DB 242 CTTATATATGCAACAGCAGACGACTTCCAGAGGTGACACCGGCTCAGCCTGCAAGTGG 183  
|||||

QY 1369 ggcctgctctatcaacagcgagcgtgtcccccagcagcagctgcgagctgagct 1428  
|||||  
DB 182 GGCTGCTCTCAATCAACAGCGGGCGCTGTCCCGACAGCAGTGGGCTGGAGCTGA 123  
|||||  
QY 1429 gactcagctgagcaga-ctcctggtgtatgtttccagaatgagcttgatgtgt 1487  
|||||  
DB 122 GACTCTAGCTGACGAGAGCTCTGCTGTATGTTTTCAGAAATGCTTGAAGTTATGTGT 63  
|||||  
QY 1488 taactctcattcgtatgctaggttatatacatgatattcaataatgaacttttaa 1547  
|||||  
DB 62 TAAATCTGCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3  
|||||  
QY 1548 ag 1549  
||  
DB 2 AG 1

RESULT 3  
AI445418/c 410 bp mRNA EST 13-APR-1999  
LOCUS t334b11.x1 NCI\_CGAP\_Pan1 Homo sapiens cDNA clone IMAGE:2143389 3',  
DEFINITION mRNA sequence.  
ACCESSION AI445418  
VERSION AI445418.1 GI:4288073  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 410)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 7, 1998 this sequence version replaced gi:3121415.  
Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Life Technologies catalog #: 11548-013  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILM at:  
[www-bio.illn1.gov/bbrp/image/image.html](http://www-bio.illn1.gov/bbrp/image/image.html)

FEATURES  
source  
Insert Length: 2579 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 368.  
Location/Qualifiers  
1..410  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2143389"  
/clone\_lib="NCI\_CGAP\_Pan1"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.72 kb. Life Technologies catalog #:  
11548-013"

BASE COUNT 100 a 99 c 91 g 120 t  
ORIGIN

Query Match 23.9%; Score 371.4; DB 46; Length 410;  
Best Local Similarity 99.5%; Pred. No. 1e-85;  
Matches 383; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1166 tccctgagagcagagcctgtctggaagaggtgagcagcgtcgcgaccagttaccacga 1225  
|||||  
DB 410 TCCTGAGAGGAGAGGCTGTGCGAAGAGGTGACCGGCTGCGGACCAAGTTACCAACA 351  
|||||  
QY 1226 tgcacaatctgattgctgacctgtgagtggaacagagtgaaataatgattacaagaaga 1285  
|||||

```
|||||
Db 350 TGCCACAATCGATTGCTGACCTGGATGGAACAGAGTGAATAATTAATTAACAAAGAGA 291
Qy 1286 tattacattcatctgtgttgacttaataatgccaacagaccagccctcccaaggt 1345
Db 290 TATTTACATTCATCTGGTGTAGACTTAATATGCAACAGCAGACCACTTCCAGAGGT 231
Qy 1346 gaacccgctcagccctgagtgaggctgtctcctcaacagcgagcgctgcgccagcag 1405
Db 230 GACACCGGCTCAGCCTGAGAGTGGGGCTGTCTCATCAACCGGGCGGTGTCCCGCAGC 171
Qy 1406 cagtcggagctgagcctgagcctgactctagctgagcaga-ctccgtgtatgttttca 1464
Db 170 CAGTCGGGCTGAGAGCTGAGTGTGACTGTAGCTGAGCAGAGCTCCTGGTATGTTTCA 111
Qy 1465 gaattgctgaagtatgttttaaatctctcattcgtatgtcaggttatataataga 1524
Db 110 GAAATGGCTTGAAGTATGTTTAAATCTGCTCATTTGCTAGGTTATACATATGA 51
Qy 1525 ttccaataatgaacttttaag 1549
Db 50 TTTTCATTAATGAACTTTAAAG 26

RESULT 4
LOCUS AI092201 391 bp mRNA EST 05-OCT-1998
DEFINITION ga60a09.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
IMAGE:1691128 3, mRNA sequence.
ACCESSION AI092201
VERSION AI092201.1 GI:3431195
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 391)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 17, 1998 this sequence version replaced gi:2045437.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 503 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 297.
Location/Qualifiers
1.391
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1691128"
/clone_1bp="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAGTGAAGGAGCGGCGCATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHH19W."
BASE COUNT 87 a 102 c 65 g 137 t
ORIGIN
```

```
Query Match 23.8%; Score 369; DB 42; Length 391;
Best Local Similarity 100.0%; Pred. NO. 4.3e-85;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 cgtgtgtgtgtcttcttgaatgtgtcagcatgaagctgtcttcccaaccat 231
Db 391 CGTGTGTGTGTCCTTTGTTGAATGTGCTCAGCATGAAGCTGTCTTCCCAACCAT 332
Qy 232 actaatgtcatatgcagaccatcgaaagcgtgttaagaagaatgactgtatgtct 291
Db 331 ACTAATGTTTCATATGACAGACCATCGAAAGACTGGTTAAAGAAAGATGACTTATGTCT 272
Qy 292 gcactagttccgfaagagagcgttgcagatgcagcgaagaagaagcaagtcatt 351
Db 271 GCAGTAGTTTCCGTAAGAGAGCAGCTTGCCAGATACGACGCAAAAGAGAGCAAGTGTAT 212
Qy 352 gaacagtgaaacaagttttgcaatatctgaggaaagccaatttgaaaaaccaagct 411
Db 211 GACAGGTGAACAAGTTTTCATAATCTGAGAGCAACCATTTTGAAAAACCAAGGCT 152
Qy 412 ttaatccagtgtagccagtgtaggaagagcctgtagagcgagcgagcgttgaaaa 471
Db 151 TTAATCCAGTGTGACCACTTGAGAGAGAGCTGAGAGCGCAGCGCACTTGAAAAA 92
Qy 472 gaattgcatctcagcaagagaagagccattgagaagaagcagatgaagaaggaata 531
Db 91 GAACTTGCTATCTCAGCAAGAGAAAAAGCCATTGAGAAAGCATGATGAAAAAGGAATA 32
Qy 532 acgaagaagaa 540
Db 31 ACGAAAGAA 23
```

```
RESULT 5
LOCUS AI266387/c 365 bp mRNA EST 29-JAN-1999
DEFINITION gm45h08.x1 Soares_placenta_8to9weeks_2NBHP8to9W Homo sapiens cDNA
IMAGE:1891743 3, mRNA sequence.
ACCESSION AI266387
VERSION AI266387.1 GI:3884545
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 365)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2282340.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1794 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 361.
Location/Qualifiers
1.365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1891743"
/clone_1bp="Soares_placenta_8to9weeks_2NBHP8to9W"
/dev_stage="two placenta: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
```



TTTACCAATCTGAAGTGGAGCGGCCGATTTTTTTTTTTTTTTT 3'),  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia) digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library constructed by Bento Soares and  
M.Fatima Bonaldo."

BASE COUNT 96 a 89 c 83 g 97 t  
ORIGIN

Query Match 22.5%; Score 349.4; DB 44; Length 365;  
Best Local Similarity 99.4%; Pred. No. 5e-80;  
Matches 361; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1191 agaggtgacggcgtcgccgagccattaccagcatgcacaacttgatctgacctg 1250  
|||||  
Db 365 AGAGGTGACCGCGTCGCGACCCAGTACCCAGCATGCGACATCTGATTCGTGACCTGG 306  
Qy 1251 atggacagagtgaaataatgaataatgaataatgaataatgaataatgaataat 1310  
|||||  
Db 305 ATGGACAGAGTGAATTAATGATTTACAAAGATATTTCATTCATCTGGTTAGACT 246  
Qy 1311 taatagccacaacgacacgacacccctccaggtgacacgacctgacgctgagtg 1370  
|||||  
Db 245 TAAATATGCCAACACGACACGACCTTCCAGAGGTGACACGCGCTCAGCTGAGGGG 186  
Qy 1371 ctgtctccatcaacgagggcgctgcccgacgacgtcgagctgagctgagctcga 1430  
|||||  
Db 185 CTGGTCCATATCAACGCGGCGCTGCCCGACGACGCGGCGGAGCTGAGCTGGA 126  
Qy 1431 ctctagctgagacaga-ctcctgtgtatgttttcagaaatgctgaaagttaatgttta 1489  
|||||  
Db 125 CTCTAGCTGAGACAGCTCTCTGTGTATGTTTTCAGAAATGCTTGAAGTTATGTTTA 66  
Qy 1490 aatctgcatctgcatgctgattacatacatgatttcaataaagaacttttaag 1549  
|||||  
Db 65 AATCTGCATCTGATGTTAGTTATACATGATGATTTTCATTAATGAACCTTTTAAAG 6  
Qy 1550 aaa 1552  
|||||  
Db 5 AAA 3

RESULT 6  
AI288955/c 360 bp mRNA EST 29-JAN-1999  
LOCUS q181c08.x1 Soares\_NHMPu\_S1 Homo sapiens cDNA clone IMAGE:1878734  
DEFINITION 3', mRNA sequence.  
ACCESSION AI288955  
VERSION AI288955  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 360)  
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Feb 11, 1998 this sequence version replaced gi:2873143.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 431 Std Error: 0.00  
Seq primer: -400P from Gibco.

FEATURES  
source  
1. .360  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1878734"

/clone.lib="Soares\_NHMPu\_S1"  
/tissue\_type="Pooled human melanocyte, fetal heart, and  
pregnant uterus"  
/lab\_host="DH10B"  
/note="Organ: mixed (see below); Vector: pT7T3D-Pac  
(Pharmacia) with a modified polylinker; Site 1: Not I;  
Site 2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NBH, pregnant uterus  
NBHPU, and fetal heart NBHRIW) were mixed, and ss circles  
were used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools  
consisted of T.M.A.G.E. clones 260232-265223,  
340488-345479, and 484488-489479."

BASE COUNT 96 a 88 c 84 g 92 t  
ORIGIN

Query Match 22.3%; Score 345.4; DB 44; Length 360;  
Best Local Similarity 99.4%; Pred. No. 5.4e-79;  
Matches 357; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1193 aggtgacggcgtcgccgagccagttaccagcatgcacaatctgattgctgacctg 1252  
|||||  
Db 360 AGGTGACCGCGCTCGCGACCCAGTACCCAGCATGCGCAATCTGATTCGACTGGAT 301  
Qy 1253 ggaacagagtgaataatgaataatgaataatgaataatgaataatgaataatga 1312  
|||||  
Db 300 GGAACAGAGTGAATTAATGATTTACAAAGATATTTCATTCATCTGGTTAGACTTA 241  
Qy 1313 aatgcccacaacgacacgacacccctccaggtgacacgacctgacgctgagtgag 1372  
|||||  
Db 240 AATATGCCAACACGACACGACCTTCCAGAGGTGACACGCGCTCAGCTGAGTGGGCT 181  
Qy 1373 ggtctcctcaacgagggcgctgtcccgacgacgtcggtgagctgagctgagct 1432  
|||||  
Db 180 GGTCTCTATCAACGCGGCGCTGTGCCGACGACGTCGGGTGGAGCTGAGCTGACT 121  
Qy 1433 ctgactgagcaga-ctcctgtgtatgttttcagaaatgctgaaagttaatgttta 1491  
|||||  
Db 120 CTAGCTGAGACAGAGCTCTCGTGTATGTTTTCAGAAATGCTTGAAGTTATGTTTAA 61  
Qy 1492 tctgctcattgctatgctgattacatacatgatttcaataaagaacttttaag 1550  
|||||  
Db 60 TCTGCTCATCTGATGTTAGTTATACATGATGATTTTCATTAATGAACCTTTTAAAG 2

RESULT 7  
AA886109 398 bp mRNA EST 30-MAR-1998  
LOCUS ny44f05.s1 NCI-CCAP\_Pri2 Homo sapiens cDNA clone IMAGE:1274625,  
DEFINITION mRNA sequence.  
ACCESSION AA886109  
VERSION AA886109  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 398)  
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 17, 1998 this sequence version replaced gi:1900538.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D.,  
Rodrigo F. Chuagui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Kitzman, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.lnl.gov/bbrp/image/image.html](http://www-bio.lnl.gov/bbrp/image/image.html)

Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 392.

## FEATURES

SOURCE

1. .398

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1274625"

/clone\_1b="NCI\_CGAP\_Pr12"

/sex="male"

/tissue\_type="metastatic prostate bone lesion"

/lab\_host="DH10B"

/note="Vector: PAMPI0; mRNA made from metastatic prostate  
lesion of the bone, cDNA made by oligo-dT priming.  
Non-directionally cloned. Size-selected on agarose gel,  
average insert size 600 bp. Library made by D. Krizman,  
NIH."

## FEATURES

SOURCE

1. .433

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1566248"

/clone\_1b="Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1"

/lab\_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
Equal amounts of plasmid DNA from five normalized  
libraries were mixed, and ss circles were made in vitro.  
Following HAP purification, this DNA was used as tracer in  
a subtractive hybridization reaction. The driver was  
PCR-amplified cDNAs from pools of 5,000 clones made from  
the same 5 libraries. The pools consisted of the following  
libraries and clones: Soares NBHSF pool 1:  
30384-310919, 323208-325895 Soares NB2HP pool 1:  
145032-147335, 147720-148103, 148872-149255, 15002 -  
150407, 151176-152327 Soares NB2HP8-9W pool 1:  
758280-760583, 772104-774407 Soares NBHPA pool 1:  
304776-306311, 320136-322823, 326280-326653 Soares NBHOR  
pool 1: 723720-726407, 739080-740999 Subtraction by Bento  
Soares and M. Fatima Bonaldo."

REFERENCE 1 (bases 1 to 384)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 14, 1998 this sequence version replaced gi:1798275.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 1616 Std Error: 0.00  
Seq primer: -40m3 fwd. ET from Amersham  
High quality sequence stop: 358.

FEATURES  
source  
1. .384  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1658236"  
/clone\_lib="Soares\_total\_fetus\_Nb2HF8\_9w"  
/dev\_stage="8-9 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from pooled 8-9 week  
(total) fetus material with a Not I - oligo(dT) primer [5'  
GTGTTACCAATCTGAGTGGAGCGCGCTTAAATTTTATTTTATTTT  
(Pharmacia), digested with Not I and Eco RI adaptors  
(Pharmacia), digested with Not I and Eco RI sites of the modified  
pT73 vector. Library is normalized, and was  
constructed by Bento Soares and M. Fatima Bonaldo. "  
BASE COUNT 114 a 83 c 82 g 104 t 1 others  
ORIGIN

Query Match 21.1% Score 327.2: DB 41: Length 384;  
Best Local Similarity 97.4%: Pred. No. 2.9e-74;  
Matches 332: Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1212 ccagttaccacacatgccacaatctgctgacctgagatggaacagagtgaaataat 1271  
|||||  
DB 342 ccagttaccacacatgccacaatctgctgacctgagatggaacagagtgaaataat 283  
|||||

QY 1272 gaattacaagaagatatattacatcctgctgctgacttaataatgcccacacacac 1331  
|||||  
DB 282 gatTTACAAAGAGATATTACATTCATCGTTTAGACTTAATATGCCAACGACACAC 223  
|||||

QY 1332 gacctcccaaggtgacacgcgcctcagctgcaagtgaggctgctcctcaacgcgggc 1391  
|||||  
DB 222 GACCTTCCAGGGGTGACACCGCTCAGCCTGCAAGTGGGCTGGTCTCATCAACCGGGC 163  
|||||

QY 1392 gctgtcccccagcagcagctgaggctgagctgagctgagctgagctgagctcctg 1451  
|||||  
DB 162 GCTGTCCCGCACGCAAGTGGGCTGAGCTGAGTGTGACTTGAAGTGCACAGCTCCTG 103  
|||||

QY 1452 gtgtatgtttccaagaatgcttgaagtatgtttaaactgctcattgctatgctag 1511  
|||||  
DB 102 GGTATGTTTTCAGAAATGCTTGAAGTATGTTTAATCTGCCTCATTCGATCTAG 43  
|||||

QY 1512 gtatacatatgattttcaataaatagaactttttaaagaa 1552  
|||||  
DB 42 GTTATACATATGATTTTCAATAAATGAACCTTTTAAAGAAA 2  
|||||

RESULT 10  
AA829894  
LOCUS AA829894 424 bp mRNA EST 07-APR-1998  
DEFINITION oes1e12.s1 NCI-CGAP\_Lus Homo sapiens cDNA clone IMAGE:1415162 3',  
mRNA sequence.  
ACCESSION AA829894  
VERSION AA829894.1 GI:2902993

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 424)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2151585.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.llnl.gov/dbtp/image/image.html](http://www.bio.llnl.gov/dbtp/image/image.html)

FEATURES  
source  
1. .424  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1415182"  
/clone\_lib="NCI-CGAP\_Lus"  
/tissue\_type="carcinoid"  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker. 1st strand cDNA was prepared from a  
neuroendocrine lung carcinoid, and was then primed with a  
Not I - oligo(dT) primer. Double-stranded cDNA was ligated  
to Eco RI adaptors (Pharmacia), digested with Not I and  
Eco RI sites of the modified pT73 vector. Library is normalized,  
constructed by Bento Soares and M. Fatima Bonaldo. "  
BASE COUNT 114 a 89 c 90 g 131 t  
ORIGIN

Query Match 21.0% Score 326.4: DB 39: Length 424;  
Best Local Similarity 99.4%: Pred. No. 4.7e-74;  
Matches 338: Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1212 ccagttaccacacatgccacaatctgctgacctgagatggaacagagtgaaataat 1271  
|||||  
DB 27 ccagttaccacacatgccacaatctgctgacctgagatggaacagagtgaaataat 86  
|||||

QY 1272 gaattacaagaagatatattacatcctgctgctgacttaataatgcccacacacac 1331  
|||||  
DB 87 gatTTACAAAGAGATATTACATTCATCGTTTAACCTTAATATGCCAACGACACAC 146  
|||||

QY 1332 gacctcccaaggtgacacgcgcctcagctgcaagtgaggctgctcctcaacgcgggc 1391  
|||||  
DB 147 GACCTTCCAGGGGTGACACCGCTCAGCCTGCAAGTGGGCTGGTCTCATCAACGGGGC 206  
|||||

QY 1392 gctgtcccccagcagcagctgaggctgagctgagctgagctgagctgagctcctc 1450  
|||||  
DB 207 GCTGTCCCGCACGCAAGTGGGCTGAGCTGAGTGTGACTTGAAGTGCACAGCTCCT 266  
|||||

QY 1451 gtgtatgtttccaagaatgcttgaagtatgtttaaactgctcattgctatgctag 1510  
|||||  
DB 267 GGTATGTTTTCAGAAATGCTTGAAGTATGTTTAATCTGCCTCATTCGATCTAG 326  
|||||

QY 1511 ggtatacatatgattttcaataaatagaactttttaaaga 1550  
|||||  
DB 327 GGTATACATATGATTTTCAATAAATGAACCTTTTAAAGA 366  
|||||

RESULT	11
LOCUS	AA879456 430 bp mRNA EST 19-MAY-1998
DEFINITION	OJ19G03.S1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:1505716 3', mRNA sequence.
ACCESSION	AA879456
VERSION	AA879456.1 GI:2988567
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 430) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2282360. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 956 Std Error: 0.00 Seq primer: -40m13 fwd. ET from Amersham High quality sequence stop: 419. Location/Qualifiers 1..430 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1505716" /clone_lib="Soares_NFL_T_GBC_S1" /lab_host="DH10B" /note="Organ: pooled; Vector: pTZ193D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCL-GAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driven was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."
FEATURES	
source	
BASE COUNT	115 a 89 c 90 g 136 t
ORIGIN	
Query Match	21.0%; Score 326.4; DB 39; Length 430;
Best Local Similarity	99.4%; Pred. No. 4.8e-74;
Matches 338; Conservative	0; Mismatches 1; Indels 1; Gaps 1;
Db	1212 ccagttaccgcgcatgcacaaatcgtattctgcacctgatgtgaacagaagttaataat 1271
	33 CCAGTTACC GCACATCTGCATTGCTGCAGCTGGATGGAACAGAGTAATAA TAAAT 92
Db	1272 gaattacaagaagatatattacatcaatcctggtttagacttaaatatgccaacagcacac 1331
	93 GATTTACAAGAAGATATTTCATCATCTGCTTGTTAGACTTAATATGCCACACGACACAC 152
Db	1332 gacctccccaagggttacaccgccttaagctctgaagcgggctgcttcataaacgccggc 1391
	153 GACCTTCCCAAGGTGTACACCCGCTTAGCCTGCAGTGGGGCTGTCTCTATCAACCGGGGC 212
Db	1392 gctgtccccgcacgcagctcgagcttgagctgtagtctgaactctaagctagacaga -ctcct 1450
	213 GCTGTCCCCGCACGCAAGTCCGGGCTGGAGCTGAGTCTGACTCTGACGACGACTCTCT 272
Db	1451 ggctgtagttttcagaagaatgcttgaagttatgvttttaaactctgcatcattcgatgcta 1510

DB	273	GGTGTATGTTTTCAGAAATGGCTTGAAGTTATGCTGTTTAAATCTGCTCATTTGTATGCTA	332
QY	1511	ggtatacatatgatattccaataatgaacttttaaga	1550
DB	333	GGTATACATATGATGATTTTCATTAATGAACCTTTTAAAGA	372
RESULT	12		
LOCUS	AA029201	439 bp mRNA	EST 09-MAY-1997
DEFINITION	ZK12108.s1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone IMAGE:470343 3', mRNA sequence.		
ACCESSION	AA029201		
VERSION	AA029201.1	GI:1496605	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 439) Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapell, B., Chisoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Madis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Roloff, T., Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J., Trevaaks, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.		
TITLE	Generation and analysis of 280,000 human expressed sequence tags		
JOURNAL	Genome Res. 6 (9), 807-828 (1996)		
MEDLINE	97044478		
COMMENT	On Sep 21, 1992 this sequence version replaced gi:279290. Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Insert Length: 2448 Std Error: 0.00 Seq primer: -40M13 fwd. from Amersham High quality sequence stop: 371. Location/Qualifiers 1. 439 /organism="Homo sapiens" /db_xref="GDB:3756345" /db_xref="taxon:9606" /clone="IMAGE:470343" /clone_lib="Soares_pregnant_uterus_NbHPU" /sex="female" /dev_stage="adult" /lab_host="DH10B" /note="Organ: uterus; Vector: pT73-Pac; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACGGAAGAATTCGCGCGCCCTTTTCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."		
BASE COUNT	120 a 92 c 91 g 136 t		
ORIGIN			
Query Match	21.0%; Score 326.4; DB 27; Length 439;		
Best Local Similarity	99.4%; Pred. No. 4.8e-74;		
Matches 338; Conservative 0; Mismatches 1; Indels 1; Gaps 1;			
QY	1212	ccagttaccagcatgcacacatctgattctgcgactgattgacacagatgaataaat	1271
DB	25	CCAGTTACCCACGACATCCACAACTGATTTGCTGACGCTGATGGAACAGAGTAATAAAT	84

[illegible]



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 19, 2000, 21:16:57 ; Search time 158.18 Seconds  
(without alignments)  
1245.804 Million cell updates/sec

Title: US-09-502-945-1  
Perfect score: 1552  
Sequence: 1 cttctgagtcacgcgagaa.....aaatgacttttaagaaga 1552

Scoring table:  
IDENTITY\_NNC  
Gapop 10.0 , Gapext 1.0

Searched: 226296 seqs, 63486255 residues

Total number of hits satisfying chosen parameters: 452592

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/PCtUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/backfilest1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	90.6	5.8	7218	1	US-08-232-463-14
2	42.4	2.7	543	7	5273901-6
3	41.8	2.7	3489	4	US-08-728-323A-1
4	41.8	2.7	32207	3	US-08-770-379-20
5	41.6	2.7	2223	2	US-08-257-073-4
6	41.2	2.7	697	7	5171843-10
7	41.2	2.7	1137	7	5171843-8
8	40.4	2.6	234	3	US-08-469-802B-3
9	40.4	2.6	234	3	US-08-267-803B-3
10	39.6	2.6	9551	4	US-08-056-200-93
11	39.6	2.6	9551	4	US-08-800-644-93
12	38.8	2.5	195	2	US-08-469-802B-2
13	38.8	2.5	195	3	US-08-267-803B-2
14	38.8	2.5	2277	2	US-08-676-967-2
15	38.8	2.5	2277	2	US-08-676-974-2
16	38.8	2.5	2277	4	US-09-098-487-2
17	38.8	2.5	4200	1	US-08-242-932-1
18	38.8	2.5	4200	2	US-08-714-481-1
19	38.8	2.5	4200	6	PCT-US95-06111-1
20	38	2.4	325	3	US-08-332-766A-11
21	38	2.4	1772	4	US-08-960-022-13
22	37.4	2.4	171	2	US-08-469-802B-5
23	37.4	2.4	171	3	US-08-267-803B-5
24	37.2	2.4	3730	1	US-08-242-932-8
25	37.2	2.4	3730	2	US-08-714-481-8
26	37.2	2.4	3730	6	PCT-US95-06111-8
27	37	2.4	1394	1	US-07-730-953-1

## ALIGNMENTS

28	36.8	2.4	168	2	US-08-469-802B-4	Sequence 4, Appl1
29	36.8	2.4	168	3	US-08-267-803B-4	Sequence 4, Appl1
30	36.8	2.4	533	7	5482709-5	Patent No. 5482709
31	36.2	2.3	43795	5	US-08-742-185-101	Sequence 101, App
32	35.8	2.3	1709	4	US-09-010-398-2	Sequence 2, Appl1
33	35.8	2.3	4363	3	US-08-685-576-5	Sequence 5, Appl1
34	35.8	2.3	7210	3	US-08-257-963B-10	Sequence 10, Appl1
35	35.8	2.3	7210	6	PCT-US95-07201-10	Sequence 10, Appl1
36	35.8	2.3	22481	6	PCT-US95-07201-43	Sequence 43, Appl1
37	35.6	2.3	2729	1	US-08-412-431-2	Sequence 2, Appl1
38	35.6	2.3	2729	5	US-08-623-679-2	Sequence 2, Appl1
39	35.6	2.3	2729	5	US-08-933-774-2	Sequence 2, Appl1
40	35.4	2.3	1647	7	5405943-3	Patent No. 5405943
41	35.2	2.3	597	3	US-08-332-766A-19	Sequence 19, Appl1
42	34.8	2.2	1535	5	US-08-747-574-3	Sequence 3, Appl1
43	34.8	2.2	3376	1	US-08-320-559-29	Sequence 29, Appl1
44	34.8	2.2	3376	6	PCT-US94-04496-29	Sequence 29, Appl1
45	34.6	2.2	154	2	US-08-469-802B-6	Sequence 6, Appl1

RESULT 1  
US-08-232-463-14/C  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZpt-Fls  
US-08-232-463-14





Db 2054 AGCAGCAGATGATGACAGCAGATGATGACAGCAGATGATGACAGCAGATG 2113  
QY 605 tggaaagtattacaagaagaattcagctcttaatacattgaggaattcaagcc 664  
Db 2114 AGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2173  
QY 665 agcttgcttcgaggaattgagtcacaagaagtgctgaggaattgagtcagctga 724  
Db 2174 AGCAGCAGATGACAGCAGCAGCAGATGACAGCAGCAGCAGCAGCAGCAGCAG 2233  
QY 725 ataaaccacactgaggaagatgagcaggaagaagcagcaggaagtcagagcaaaa 784  
Db 2234 ATGAGCAGCAGCAGCAGATGAAACAGAGCAGCAGCAGCAGCAGCAGCAGCAGC 2293  
QY 785 ctaacaggatcttgaattaaagatcaggaataagagaattgagaataagaaatgagtc 844  
Db 2294 AGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2353  
QY 845 aaagcaacaacactgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 904  
Db 2354 AGCAGCAGCAGCAGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2413  
QY 905 tgaactaacagaaactgctgagcgaattctgagcaccactgcaacctcaccagatcga 964  
Db 2414 AGCAGCAGCAGCAGGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2473  
QY 965 tagctcaactcagtcagaagaagaagatatacatatgataattggaagaaatcagaga 1024  
Db 2474 TAGAGCAGCAGCAGCAGGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2533  
QY 1025 gaaatgaagaattgaggaagcagtgctgagcagcagcagcagcagcagcagcagcagc 1084  
Db 2534 AAGAGCAGCAGCAGCAGGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2593  
QY 1085 aaaggtcaagcagcagtcgtaagcagcagcagcagcagcagcagcagcagcagcagc 1144  
Db 2594 AGGTGAGCAGCAGCAGCAGGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2653  
QY 1145 tcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1197  
Db 2654 AAGAGCAGCAGCAGCAGGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2706

RESULT 4  
US-08-770-379-20/c  
; Sequence 20, Application US/08770379  
; Patent No. 5849564  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Yuan  
; APPLICANT: Bohenzky, Roy A.  
; APPLICANT: Russo, James J.  
; APPLICANT: Edelman, Isidore S.  
; APPLICANT: Moore, Patrick S.  
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED  
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/770.379  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 52342  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32207 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-770-379-20

Query Match 2.7%; Score 41.8; DB 3; Length 32207;  
Best Local Similarity 40.9%; Pred. No. 0.18;  
Matches 316; Conservative 0; Mismatches 457; Indels 0; Gaps 0;  
QY 425 accagttgaggaagcagtcgaggaagcagcagcagcagcagcagcagcagcagcagc 484  
Db 20063 AGCAGCAGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 20004  
QY 485 agcaagaagaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 544  
Db 20003 AGCAGCAGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19944  
QY 545 agtacttgatcaagaatgagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 604  
Db 19943 AGCAGCAGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19884  
QY 605 tggaaagtattacaagaagaattcagctcttaatacattgaggaattcaagcc 664  
Db 19883 AGCAGCAGCAGCAGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 19824  
QY 665 agcttgcttcgaggaattgagtcacaagaagtgctgaggaagaaatgagtcagctga 724  
Db 19823 AGCAGCAGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19764  
QY 725 ataaaccacactgaggaagatgagcaggaagaagcagcaggaagtcagagcaaaa 784  
Db 19763 ATGAGCAGCAGCAGCAGGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19704  
QY 785 ctaacaggaattctgaattaaagatcaggaataagagaattgagaatagaactgagtc 844  
Db 19703 AGCAGCAGCAGGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19644  
QY 845 aaagcaacaacactgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 904  
Db 19643 AGCAGCAGCAGCAGGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19584  
QY 905 tgaactaacagaaactgctgagcgaattctgagcagcagcagcagcagcagcagcagc 964  
Db 19583 AGCAGCAGCAGCAGGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19524  
QY 965 tagctcaactcagtcagaagaagaagatatacatatgataattggaagaaatcagaga 1024  
Db 19523 TAGAGCAGCAGCAGGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19464  
QY 1025 gaaatgaagaattgaggaagcagtgctcagcagcagcagcagcagcagcagcagcagc 1084  
Db 19463 AAGAGCAGCAGCAGGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19404  
QY 1085 aaaggtcaagcagcagtcgtaagcagcagcagcagcagcagcagcagcagcagcagc 1144  
Db 19403 AGGTGAGCAGCAGCAGGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19344  
QY 1145 tcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1197  
Db 19343 AAGAGCAGCAGCAGGATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19291

```
RESULT 5
US-08-257-073-4
; Sequence 4, Application US/08257073
; Patent No. 576597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Curtiss, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-257-073-4

Query Match      2.7%; Score 41.6; DB 2; Length 2223;
Best Local Similarity 56.6%; Pred. No. 0.045;
Matches 77; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
```

```
; TITLE OF INVENTION: IMMUNOGENIC POLYPEPTIDE AND METHOD FOR
; PURIFYING IT
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/175,112
; FILING DATE: 30-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 754,645
; FILING DATE: 9-JUL-1985
; APPLICATION NUMBER: 115,634
; FILING DATE: 26-OCT-1987
; APPLICATION NUMBER: 649,903
; FILING DATE: 12-SEP-1984
; SEQ ID NO: 10:
; LENGTH: 697
; 517843-10

Query Match      2.7%; Score 41.2; DB 7; Length 697;
Best Local Similarity 44.1%; Pred. No. 0.031;
Matches 267; Conservative 0; Mismatches 333; Indels 6; Gaps 2;
```

```

;
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/175,112
; FILING DATE: 30-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 754,645
; FILING DATE: 9-JUL-1985
; APPLICATION NUMBER: 115,634
; FILING DATE: 26-OCT-1987
; APPLICATION NUMBER: 649,903
; FILING DATE: 12-SEP-1984
; SEQ ID NO: 8
; LENGTH: 1137
5171843-8

```

```

Query Match      2.7%; Score 41.2; DB 7; Length 1137;
Best Local Similarity 44.1%; Pred. No. 0.04;
Matches 267; Conservative 0; Mismatches 333; Indels 6; Gaps 2;

```

```

QY 601 caggtggaagaaagttacaaagaaagatttcagctattatcaactgaggaattcaa 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 248 cagaacccaataatccacgtgaataaagctgaagcaacagagagacagatgagac 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 661 agccagctgcttcgagaaatgagatgcacaaagtgctggaagaaatgcgtatcag 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 308 agccagcagagagacagagatgagacagcagcaggtgatagacagatgagcaaccag 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 721 ctgaataaaacaaactgagag--aggaatgagcagaaagagacagagatcaga 777
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 368 cagagatagagcagctgagacacacagcagagatagcagatgagcagcagcagagag 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 778 gcaaaaaactaaagagatcttaaatgaatcaggaatagagaatgaggaatagaga 837
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 428 acagagcagatgagacagcagagagagacagagcagatgagacacagcagagagagag 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 838 ctgagatgaagaaacaaacactctggaacagagcagcagcagcagcagcagcagagag 897
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 488 cagatgagacacacacagctgagatagacagctgagacacacagctgagatagacagctg 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 898 gagtgctgagactaacaagacttgctggaatctgagacacaaactgacactcaccaga 957
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 548 gacacacagcagagatagagcagatgagcagcagcagcagagatagagagcagctgagcagc 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 958 tctggaatagctcaactcagctcagtaaaaaaaggtatcatatgataaatgaggaagtta 1017
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 608 cagcagagagatagacagatgagacagcagcagcagagatagacagctgagacagcagcag 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1018 cagagaaagaatgaaagatgagagaaacagctgtgtccacatgagagagatcatatgagacg 1077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 668 gagatagagcagatgagacacagcagcagagagatagacagctgagacagcagcagaga--g 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1078 atgaagcaaaagctlaagcagctgagataagcacaagcacaagcacaagcagcagctgag 1137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 725 atagagcagctgagacagcagcagcagagatagagcagctgagacagcagcagagagatagag 784
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1138 cagctcctcagcagcagacacagcagcttctctgagagagcagagcctgtctggaagagctg 1197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 785 cagctgagacagcagcagcaggaatgtgctgagcagcagcagcagcagagagaaacgcagcagag 844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1198 gacgcg 1203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 845 gacagcag 850
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 8
US-08-469-802B-3
; Sequence 3, Application US/08469802B
; Patent No. 5741645
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Rannum, Laura P. W.
; APPLICANT: Chung, Ming-yi

```

```

; APPLICANT: zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mueling, Raasch, Gebhardt & Schwappach, P.A.
; STREET: 119 No. 5741645th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,802B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueling, Ann M.
; REGISTRATION NUMBER: 33, 977
; REFERENCE/DOCKET NUMBER: 110.00030101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1225
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-469-802B-3

```

```

Query Match      2.6%; Score 40.4; DB 2; Length 234;
Best Local Similarity 50.5%; Pred. No. 0.028;
Matches 98; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

```

```

QY 1038 ggaagacagctgtgtccacatgagagagatcatatgagacatgagaaagcagtaagca 1097
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 30 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 89
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1098 gctgataagcacaagcacaagcacaagcagcagctgtgtcagctcctcagcaagcagaa 1157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 90 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1158 cagcttctctggaagcagagcagcagctgtcgaagaagtgaaacgagctgagacagatt 1217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 150 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1218 acccagatgcac 1231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 210 CCTCAGCAGCGCTC 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 9
US-08-267-803B-3
; Sequence 3, Application US/08267803B
; Patent No. 5834183
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Rannum, Laura P. W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Mueling, Raasch, Gebhardt & Schwappach, P.A.  
STREET: P. O. Box 581415  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55458-1415  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/267,803B  
FILING DATE: 28-JUN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: McCormack, Myra H.  
REGISTRATION NUMBER: 36,602  
REFERENCE/DOCKET NUMBER: 110.00030120  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-305-1217  
TELEFAX: 612-305-1228  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 234 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-267-803B-3

Query Match	2.6%	Score	40.4	DB	3	Length	234
Best Local	Similarly	50.5%					
Matches	98	Conservative	0	Mismatches	96	Indels	0
						Gaps	0

[illegible]

RESULT 10  
 US-08-056-200-93  
 : Sequence 93, Application US/08056200  
 : Patent No. 5615500  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Steinert, Peter M.  
 :  
 : APPLICANT: Lee, Seung-Chul  
 :  
 : APPLICANT: Kim, In-Gyu  
 :  
 : APPLICANT: Chung, Soo-Il  
 :  
 : TITLE OF INVENTION: Park, Sang-Chul  
 :  
 : TITLE OF INVENTION: Trichopyalin and Transglutaminase-3 and  
 :  
 : NUMBER OF SEQUENCES: 117  
 :  
 : CORRESPONDENCE ADDRESS:  
 :  
 : ADDRESSEE: Knobbe, Martens, Olson & Bear  
 :  
 : STREET: 620 Newport Center Drive, Sixteenth Floor  
 :  
 : CITY: Newport Beach  
 :  
 : STATE: CA  
 :  
 : COUNTRY: U.S.A.  
 :  
 : ZIP: 92660

```

1 COMPUTER READABLE FORM:
2 MEDIUM TYPE: Floppy disk
3
4 COMPUTER: IBM PC compatible
5 OPERATING SYSTEM: PC-DOS/MS-DOS
6 SOFTWARE: PatentIn Release #1.0, Version #1.25
7
8 CURRENT APPLICATION DATA:
9 APPLICATION NUMBER: US/08/056,200
10 FILING DATE: 30-APR-1993
11 CLASSIFICATION: 435
12
13 ATTORNEY/AGENT INFORMATION:
14 NAME: Fredrick, Michael F.
15 REGISTRATION NUMBER: 36,799
16 REFERENCE/DOCKET NUMBER: NIH054,001A
17
18 TELECOMMUNICATION INFORMATION:
19 TELEPHONE: (714) 760-0404
20 TELEFAX: (714) 760-9502
21
22 INFORMATION FOR SEQ ID NO: 93:
23
24 SEQUENCE CHARACTERISTICS:
25
26 LENGTH: 9551 base pairs
27
28 TYPE: nucleic acid
29 STRANDEDNESS: single
30 TOPOLOGY: linear
31 MOLECULE TYPE: cdna
32
33 HYPOTHETICAL: NO
34
35 ANTI-SENSE: NO
36
37 FEATURE:
38
39 NAME/KEY: CDS
40 LOCATION: 1507..1644
41
42 FEATURE:
43
44 NAME/KEY: Intron
45 LOCATION: 1645..2511
46
47 FEATURE:
48
49 NAME/KEY: CDS
50 LOCATION: 2512..8070
51
52
53 US-08-056-200-93

```

[illegible]

RESULT 11  
 US-08-800-644-93  
 ; Sequence 93, Application US/08800644  
 ; Patent No. 5958752  
 ;  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Steinert, Peter M.  
 ; APPLICANT: Lee, Seung-Chul  
 ;  
 ; APPLICANT: Kim, In-Gyu  
 ; APPLICANT: Chung, Soo-Il  
 ; APPLICANT: Park, Sang-Chul  
 ;  
 ; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
 ; TITLE OF INVENTION: Methods of Using Same  
 ;  
 ; NUMBER OF SEQUENCES: 117  
 ;  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear



RESULT 14  
US-08-676-967-2  
Sequence 2, Application US/086766967  
Patent No. 5747317  
GENERAL INFORMATION:  
APPLICANT: COLLINS, KATHLEEN  
TITLE OF INVENTION: Human Telomerase  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Science & Technology Law Group  
STREET: 268 Bush Street, Suite 3200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/676,967  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard A  
REGISTRATION NUMBER: 36,627

RESULT 15  
 US-08-676-974-2  
 : Sequence 2, Application US/08676974  
 : Patent No. 5770422  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: COLLINS, KATHLEEN  
 : TITLE OF INVENTION: Human Telomerase  
 : NUMBER OF SEQUENCES: 10  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESS: Science & Technology Law Group  
 : STREET: 268 Bush Street, Suite 3200  
 : CITY: San Francisco  
 : STATE: CA  
 : COUNTRY: USA  
 : ZIP: 94104  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patentin Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/676,974  
 : FILING DATE:  
 : CLASSIFICATION: 530  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Osman Ph.D., Richard A

? REGISTRATION NUMBER: 36, 627  
 ? REFERENCE/DOCKET NUMBER: OCB96-055  
 ? TELECOMMUNICATION INFORMATION:  
 ? TELEPHONE: (415)343-4341  
 ? TELEFAX: (415)343-442  
 ? INFORMATION FOR SEQ ID NO: 2:  
 ? SEQUENCE CHARACTERISTICS:  
 ? LENGTH: 2277 base pairs  
 ? TYPE: nucleic acid  
 ? STRANDEDNESS: double  
 ? TOPOLOGY: linear  
 ? MOLECULE TYPE: cDNA  
 ? OS-08-676-974-2

Query Match	2.5%	Score 38.8;	DB 2;	Length 2277;
Best Local Similarity	25.6%	Pred. No. 0.28;		
Matches 99;	Conservative 72;	Mismatches 216;	Indels 0;	Gaps 0

[illegible]

Search completed: May 20, 2000, 00:27:48  
Job time: 11451 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 19, 2000, 21:19:08 ; Search time 243.1 Seconds

(without alignments) 1597.277 Million cell updates/sec

File: US-09-502-945-1

Perfect score: 1552

Sequence: 1 cttctgagtcacgcagagaa.....aaatgaacttttaagaaga 1552

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1305	84.1	1789	1 X40059	Colon cancer assoc
2	1277.4	82.3	2353	1 V40542	Homo sapiens sec
3	852	54.9	852	1 X40054	Colon cancer assoc
4	279.4	18.0	310	1 V89187	EST clone CH699. N
5	114.4	7.4	317	1 T24815	Human gene signatu
6	113.8	7.3	432	1 X51883	Human secreted pro
7	67.2	4.3	1686	1 Q87587	DNA encoding leuco
8	57.2	3.7	4000	1 T91902	Mannose-1-phosphat
9	52.2	3.4	1435	1 X32489	Seg ID No:4 used i
10	52.2	3.4	2100	1 V42601	Streptococcus uber
11	44.8	2.9	1908	1 N71064	Gene encoding Plas
12	44.8	2.9	6755	1 V21511	Staphylococcal bac
13	44.2	2.8	3399	1 T05868	Chicken leucocytos
14	44.2	2.8	3586	1 T46187	DNA encoding cell
15	44.2	2.8	3624	1 T46186	DNA encoding cell
16	43.8	2.8	2640	1 V04529	Human OC-116 kDa c
17	43.2	2.8	1048	1 X32486	Seq ID No:1 used i
18	42.4	2.7	11000	1 V21209_05	Continuation (6 of
19	42.4	2.7	543	1 Q23092	Antigen tc-7a gene
20	41.8	2.7	3393	1 X24681	Human synaptonemal
21	41.8	2.7	202	1 T23845	Human gene signatu
22	41.8	2.7	33207	1 V73805	KSHV LTR DNA (nucl
23	41.8	2.7	137507	1 V19941	KSHV long unique c
24	41.6	2.7	932	1 V35340	DNA encoding a PII
25	41.6	2.7	2223	1 Q80908	Plasmodium falcipa
26	41.2	2.6	1137	1 Q33061	Plasmodium vivax c
27	41.1	2.6	724	1 X39663	Renal cancer assoc
28	41.1	2.6	862	1 X39858	Gastric cancer ass
29	40.6	2.6	2010	1 N91379	Intron 1 from huma
30	40.4	2.6	234	1 Q84832	Spinocebellar at
31	40.2	2.6	700	1 V88907	EST clone IA300. N
32	40.2	2.6	2223	1 Q29187	cDNA encoding Plas
33	39.2	2.5	1298	1 X40056	Colon cancer assoc
34	39.2	2.5	2277	1 V13834	Homo sapiens ambig

35	39.2	2.5	3653	1 X20280	Borrelia burgdorfe
36	39	2.5	6644	1 X33181	Base sequence of t
37	39	2.5	7372	1 X33182	Base sequence of t
38	39	2.5	7797	1 X33180	Cowpox virus bsr f
39	39	2.5	7996	1 X33184	Base sequence of t
40	38.8	2.5	195	1 Q84831	Spinocebellar at
41	38.8	2.5	1397	1 T84935	Human prostate pro
42	38.8	2.5	1797	1 T84931	Human telomerase p
43	38.8	2.5	2277	1 V05370	DNA encoding C-Det
44	38.8	2.5	3492	1 V11348	Group B Streptococ
45	38.8	2.5	4200	1 T03190	

## ALIGNMENTS

RESULT 1  
ID X40059 standard; DNA; 1789 BP.  
AC X40059.

DT 02-JUL-1999 (first entry)  
DE Colon cancer associated gene.  
KW Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KW prostate cancer; ss.  
OS Homo sapiens.  
PM W09504265-82.  
PD 28-JAN-1999.  
PF 15-JUL-1998; U14679.  
PR 22-JUN-1998; US-102322.  
PR 17-JUL-1997; US-896164.  
PR 10-OCT-1997; US-061589.  
PR 10-OCT-1997; US-061765.  
PR 11-OCT-1997; US-948705.  
PR 11-OCT-1997; GB-021697.  
PA (LUDWIG-) LUDWIG INST CANCER RES.  
PI Chen Y, Gout I, Gure A, O'hare M, Obara Y, Old LJ,  
PI Pfundschuh M, Sahn U, Scanlan MJ, Stockert E,  
PI Tureci O;  
DR WPI: 99-132448/11.

PT New isolated cancer associated nucleic acids and polypeptides -  
PT Isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers

PS Claim 67; Page 660; 787pp; English.

CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
SQ Sequence 1789 BP; 622 A; 341 C; 452 G; 374 T;

Query Match 84.1%; Score 1305; DB 1; Length 1789;  
Best Local Similarity 86.8%; Pred. No. 0;  
Matches 1552; Conservative 0; Mismatches 0; Indels 237; Gaps 1;

QY	1	cttctgagtcacgcagagaaacttacttataaggaagtgtaaatgagga 60
DB	1	CTTCTGAGTCACGCCAGAACTTAATCTTATGAGAGAAAGTGAAATGAGGA 60
QY	61	tccaatgaagtcttttggaagcagctagctgaatcatcagagaactgtgaagctt 120
DB	61	TCCCAATTGAAGTTTGTGAGAAAGCACTTACCTGCAATATCAGAGAACTGTGAAGATCTT 120

Oy	121	aaagagcaactaaagcgtlaaagaattctcttgctgcgtcaactctgttaacccggttgct	180
Db	121	AAAGAGCAACTAAAGCAATAAAGAAATTTCTTGGCGCTTAATACCTTGACCGTGTGTG	180
Oy	181	ggtcttggttgaatgtgtctcaagctgaagctggtctcttcccaaccatactaatt	240
Db	181	GGCTTTGTTGAAATGTGCTCGACATGAAGCTGTTCTTCCCAACCATACATATGTT	240
Oy	241	catactgagaccatcgaanaactcgtgttlaaagaagaatgactgtgtctgcaactagt	300
Db	241	CATATCGAGACATCGAANAACCTGTTTAAAGAAAGAGATGATCTGATCTGCACTAGTT	300
Oy	301	tcctgaaggagcagcttgtagagataagccagaaaggaagcaagtgcttatgaacaagt	360
Db	301	TCCCTAAGGAGCACCTTGGCATGCGACGACAAAGGAAGCAATGCTTATGAAACGGTG	360
Oy	361	aaacaagttctgcaatatctgtagaagccaatcttggaaaacaacaagctttaaccag	420
Db	361	AAACAAGTTTTCGAAATATCTGAGGAAGCCAAATTTGAAAAAACAGGCTTTAATCCAG	420
Oy	421	tgtagcagcttgtagaagagcctgtagaagcgagcgagcgtgacattgaaaagaactgca	480
Db	421	TGTGACCGTTGAGAGGAAGGAGCGTGGAGGCGAGCGCGACTTGAAAAAGAACTTGCA	480
Oy	481	tctcaagaagagaaagcgccattgagaagaacatgatagtaaaaaagaatacgaagaag	540
Db	481	TCTCAGCAAGAGAAAGGCGCATTTGGAAGAAAGCATGATGATAAAGAAATATACGMAA	540
Oy	541	aggagatcacatgagatcaaaagatgttgatctgtctccagaatatgtgccactgtagcc	600
Db	541	AGGAGATCATGGGATCAAAAGATGTTGATCTGTCTCAGAAATATTTGCCCACTGGAGGCC	600
Oy	601	caggctggaagaaagftlacaagaagaaagatltcagctatatacacaactgtaggaattcaa	660
Db	601	CAGGTGGAAGAAAGTTTACAAAGAAAGAAATTTTCAGCTTAATACACTGGAGAAATTCAA	660
Oy	661	agccagctggtctctcgggaatgtagatgtaacaaagtgtgtgtagaataatgcgtatacg	720
Db	661	AGCCAGCTGGCTTCTCGGGAATGAGATGTGCACAAAGTGTGGGAAATGGCGCTATACG	720
Oy	721	ctgataaacaacaacatcgtagaagagatgtagcgcagaagaagacacagaagatctcagaca	780
Db	721	CTGAAATTAACCAACATGAGGAAGAGATGAGGCCAGAAAGAGACACAGAGATTCAGAGCA	780
Oy	781	aaaactacagggatctctgaaattlaaagatcagagaatatgagaanaaltgagataagaactg	840
Db	781	AAACTCAACAGGGATCTTTGAATTTAAAGATCAGGAATAGGAATTTGGAATGGAACGTG	840
Oy	841	gattgaagcaacaacaactctgagaacagagcagcagaagcgacccctgcagagagagag	900
Db	841	GATGAAGCAACAACACTTGTGAACAGGAGCAGACAGCAAGGCCCTGTGGCCAGAGAGAG	900
Oy	901	tgcttgagacttaacagaaactgctgtagcgatctgtagcaccactgtcactc-----	952
Db	901	TGCTTGAGACTTAACAGAACTGTTGGCGGATGTGAGCAACCACTGCACCTCACAGACAG	960
Oy	952	-----	952
Db	961	GAAGAAATGACATTTAGCGAGACTTTAGCAAGGAAGCAAGGCCCAAGCCCTTCAGGCC	1020
Oy	952	-----	952
Db	1021	CAGCAAGAGAGCGAGAGCTGACACAGAGATACAGCAAAATGGAAGCCAGCATACAAA	1080
Oy	952	-----	952
Db	1081	ACTGAATGACAGATTTGTTGCTGACCTCCAGAAATACATTTTGGACAAAGTTAAAG	1140
Oy	952	-----	952
Db	1141	GAACAATGCTGTACATTAGCCAAAGAAACGTGGAACAAATCTCTCAAAAAAACAGATCTGAA	1200
Oy	964	atagctcaactcagttcaagaanaaaggtlatalatgataaatttggaagaagtltcaagaga	10233

Db	1201	ATAGCTCACTCACTCAAGAAAAAGGTATACATATGATTAATTTGGAGAAAGTTACAGACA	1260
Qy	1024	agaaatgaagatcttggagagacagctgtctcaacatctggagatcatatgaacgaatgag	1083
Db	1261	AGAAATGAAGAATTTGGAGAAACAGTGTGTCTCCACATGGGAGAGTACATGACGATGAAG	1320
Qy	1084	caaaagctaaagcagcctgtgataagcacagccagccacagcccaagcagctgtgtcagctc	1143
Db	1321	CAAAAGGCTTAAGCGCTGTGGATTAAGCACAGCCAGCCACAGCCACAGCTGTGTGCAGCTC	1380
Qy	1144	ctcagcaagcagcaacacagctctccctctggagaagcagaagcctctctggaagaagtggacccg	1203
Db	1381	CTCAGCAGACGAAACAGCTTCTCTCTGGAAAGGCAAGCCTGTCCGAAAGGTGGACCCG	1440
Qy	1204	ctcgagaccctgtatcccaacgacatgcacaaatctgatatctgtgacctgagatgacagagtg	1263
Db	1441	CTCGGAGCCAGATTACCCACATGCCACATATCGATGTCTGACACTGAGATGAGAAAGAGACTG	1500
Qy	1264	aaataaatgaattacaaagaagatattacatcattctgtgtttagaactaataatgccaaca	1323
Db	1501	AAATAAATGAATTTCAAAAGAGATATTACATTCATCTGCTTAGACTTAAATGCCACAA	1560
Qy	1324	cgcaccaagaccttcccaaggttgacacccgctcagcttgcaatggggctgtgtccctacata	1383
Db	1561	CGCACACAGACCTTCCACAGGCTGACACCCGCTTCAGCTCAGCTGAGGGGCTGGCTCTCATCA	1620
Qy	1384	acgagggagcctgtcccccagcagcagctgagctgagctgagatctgactctagctgaagca	1443
Db	1621	ACGGGGGCGCTGTCCCGCACAGCATGTCGGCTGGAGCTGGAGCTGACACTGACATGCGACACA	1680
Qy	1444	gactcctgtgtatgtgtttcagaagaatgagcttgaagttatgtgtttaaactctgctcatcg	1503
Db	1681	GACTCTCTGCTGTATGTTTTCAGAAATGGCTTGAAGTTATGTGTTAAATCTGCTCATTCG	1740
Qy	1504	tatctgtggtatcatcatatgattttcaataaagaacttttaagaagaa	1552
Db	1741	TATGCTAGGTATATACATATGATTTTCAATTAAGAACTTTTAAAGAA	1789
RESULT 2			
V40542			
AC	V40542	standard; cDNA; 2353 BP.	
DT	09-NOV-1998	(first entry)	
DE	Homo sapiens	secreted protein clone CH699_1.	
KW	Clone; secreted protein; ds.		
OS	Homo sapiens.		
FX	Key	Location/Qualifiers	
FT	CDS	186..203	
FT		/*tag="a	
FT		/note="secreted protein"	
PN	W09832853-A2.		
PD	30-JUL-1998.		
PF	23-JAN-1998; U01396.		
PR	24-JAN-1997; US-788789.		
PA	(GENY) GENETICS INST INC.		
PI	Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,		
PI	Racie LA, Spaulding V, Treacy M.		
DR	WPI: 98-427949/36.		
DR	P-PSDB: W29672.		
PT	New isolated polynucleotide(s) and secreted proteins - isolated from		
PT	human foetal kidney, adult brain, adult salivary gland, foetal brain		
PT	and adult testes cDNA libraries		
PS	Claim 22: Page 68-69; 109pp; English.		
CC	The sequence is that of encoding a secreted protein. Such a		
CC	protein can have biological activities, e.g. nutritional		
CC	activity, cytokine and cell proliferation/differentiation activity,		
CC	immune stimulating or suppressing activity, haematopoiesis regulating		
CC	activity, tissue growth activity, activin/inhbin activity,		
CC	chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,		
CC	receptor/ligand activity, anti-inflammatory activity, cadherin/tumour		
CC	invasion/suppressor activity, tumour inhibition activity, and other		

CC activities.  
Sequence 2353 BP; 826 A; 451 C; 580 G; 496 T;  
Query Match 82.3%; Score 1277.4; DB 1; Length 2353;  
Best Local Similarity 86.5%; Pred. No. 0;  
Matches 1536; Conservative 0; Mismatches 1; Indels 238; Gaps 2;  
QY 16 gagaagcctaaacttacttacttgaagaaagtgtgaattggaatcccaattgaattc 75  
DB 557 GAGAGCTGAAAACCTTCACTTATGAGAAAAAGTGTGAATTGAGAAATCCCAATTGAAAGTTT 626  
QY 76 ttgaagacagacttgaagatcagagaacttgaagatcttaagaagaagaactaaag 135  
DB 627 TTGAGGAACGACTTATGCTGAATATCAGAGAACTTGTGAAGATCTTAAAGACAACTTAAG 686  
QY 136 cataaagaattctctgtgctgtaacttgaacggtgtgtgtgcttctgtttgaa 195  
DB 687 CATAAAGAAATTTCTTGTGCTGCTAACTGTAAACCGTGTGGTGTCTTTGTTGAAA 746  
QY 196 tgtgtcagcatgaagcgtgtcttcccaaccaactaatgttcatatgcagccatc 255  
DB 747 TGTGCTCAGCATGAAGCTGTTCTTCCAAACCCATTAATGTTCAATGCAAGCAACATC 806  
QY 256 gaaagactggttaagaagaagaatgactgagtgtgtgactgaatttcgttaagagagc 315  
DB 807 GAAAGACTGCTTAAAGAAAGAGATGACTTGTATGCTGCACTAGTTTCCGTAAGAGCAGC 866  
QY 316 ttgacagaiaacagcaagaagaagaagtgcttaagaagtgtaaaacaagttttgca 375  
DB 867 TTGGCAGATACGACAGCAAGAAAGCAAGTGTATGAACAGGTGAAACAAGTTTGGCA 926  
QY 376 atatctgaggaagccaatttgaaaaaaaccaagcttaatccagtgfgaccagttgag 435  
DB 927 ATATCTGAGGAGCCAAATTTTGAAAAAACCAAGGCTTTAATCCACTGTGACCAAGTTGAGG 986  
QY 436 aaggaagctggaagagcgagcgagagacttgaagaagaactgcatctcagcaagaagaa 495  
DB 987 AAGGAGCTGAGAGGCGAGCGAGCGACTTGAAAAAGAACTTGCACTCAGCAAGAAAGAA 1046  
QY 496 agggccatctgaagaagaatgatgaagaagaatacaacgaagaagaaggtactgga 555  
DB 1047 AGGGCATTTGAGAAAGATGATGAAAAAGAAATACGAAGAAAGGAGATACATGGCA 1106  
QY 556 tcaaatgtgtgacttctgtctcagaataatgcccactggaagcccaagtggaagaagtlc 615  
DB 1107 TCAAGAGATGTTGATCTGTTCTCAGAAATATGCCCCAACTGAGGCCCCAGGTGAAAAAGTT 1166  
QY 616 acaaaagaaaagaatttcgctatataactgaactggaagaattcaaaagcagctgtcttc 675  
DB 1167 ACAAAGGAAAAAGATTTCACTATTATCAACTGAGGAAATTCAAAGCCAGCTGGCTTCT 1226  
QY 676 cgggaagaatgagatgcacaaaggtgtgtgagaagaatgcgctatcaactgaataaaaccac 735  
DB 1227 CGGGAATGAGATGACAAAGGTGTGTGAGAAATGCCCTACACTGATATTAACCAAC 1286  
QY 736 atggaagaagatgagcagaagaagagcagagaagttcagaagcaaaaaactaacaggaat 795  
DB 1287 ATGGAGAGAGATGAGGCGAAAAAGAGCACAGAGAGTTCAAGCAAAAAATCAACGGGAT 1346  
QY 796 ctgaagaatgaagatcaggaagaatgagaatgtgagaatagacttgatgaagaagcaaaa 855  
DB 1347 CTGGAATTTAAAGATCAGGAAATAGAGAAATTTGAGAAATGAACTGAAAGCAAAACAA 1406  
QY 856 caacttgacagagagcagcagaagcagccctgtgcagagagagatgtgctgagactaa 915  
DB 1407 CACTTGGAAACAGAGACACAGAGCAACCCCTGGCCAGAGAGAGATGCTGTGAGATAACA 1466  
QY 916 gaactgtctggcgaaatctgagcaccactgcactc----- 952  
DB 1467 GAAGTGTGGCGGAGATGTGAGCACCACTGCACCTCACACAGACAGAAAAAGATAGCATTT 1526  
QY 952 ----- 952

DB 1527 CAGCAGAGCTTTAGCAGAGAACCAAGGCCCAAGCCCTTCAGGCCCCAGCAAGAGAGCAG 1586  
QY 952 ----- 952  
DB 1587 GAGCTGACACAGAGAGATACAGCAAAATGGAAGCCAGCATGACAAAATGTGAATGACAG 1646  
QY 952 ----- 952  
DB 1647 TATTTGTTGCTGACCTCCAGAAATACATTTTGTACAAAAGTTAAAGAAAGATGCTGACA 1706  
QY 952 -----accagatctgaatgactgaactgaat 978  
DB 1707 TTAGCCAGAAACTGGAACAAATCTCTCAAAAACACAGATCTGAAATAGCTTAACCTAGT 1766  
QY 979 caagaaaaaagtlacalatalgataaattggygaagatllacagagaagaatgaagaatgt 1038  
DB 1767 CAAGAAAAAGGTATACATATGATTAATTTGGGAAGTTTACAGAGAAGAAATGAGAATTTG 1826  
QY 1039 gaggaacagtggtgtccagcatgtggaagatgacatggaagatgaagcaaaagctaaagcag 1098  
DB 1827 GAGGAACAGTGTGTCCAGCATGGAGAGTACATGAGAGATGAACCAAGGCTTAAGGCGAG 1886  
QY 1099 ctggaataagcaagccagccagccagccagctggtgtgcaagctcctcagcaagcagagac 1158  
DB 1887 CTGGAATAGCAGACAGCCAGGCCACAGCCAGCAGCTGTGTACACTCTCAGCAAGCAGAAC 1946  
QY 1159 cagcttctccttggaagagcagaagcctgtgcgaagagtgtagccagctgtcgagccagtla 1218  
DB 1947 CAGCTTTCCTGAGAGAGGAGAGGCTGTGGAAGAGGTGAGAGCGGCTGGGAGCCAGTTA 2006  
QY 1219 cccagcatgcccacatctgattgtctgacctggaatggaacagagtgaaataatgaattac 1278  
DB 2007 CCGACATGCGCAATCTGATTGTGACCTGTGATGGAACAGAGTGAATAATGATTTAC 2066  
QY 1279 aagagatatattacatctcgtgtgttgaacttaatgatcccaacgcaaccagcattc 1338  
DB 2067 AAGAGATATTTTACATTCATCTGTGTTTACTTAAATATGCCCACCAAGCCAGCATTC 2126  
QY 1339 ccaggtgtacacccgctcagcctgagctgtgagctgtgctcctcaacagcgagcgtgtcc 1398  
DB 2127 CCAGGCTGACACCGGCTCAGCCTGAGCTGAGTGGGCTGTCTCATCAACGGGGCGCTGTCC 2186  
QY 1399 ccgcagcagctcgtgagctgtgagctgtgagctgtgacttgaactgagcaga-ctcctgtgtat 1457  
DB 2187 CCGCAGCAGAGCTGGGCTGTGAGCTGTGAGTCTGACTAGCTGAGCAGAGCTCCTGCTGTAT 2246  
QY 1458 gtttcagaagaatggttgaagttatgtgtttaaactctgaatctgtatgtcagttata 1517  
DB 2247 GTTTTCAGAAATGGCTTGAAGTTATGTGTTTAAATCTGCTCAATGATGCTAGGTTATA 2306  
QY 1518 catatgatttcaataaataatgaacttttaaagaa 1552  
DB 2307 CATATGATTTTCAATTAATGAACCTTTTAAAGAA 2341  
RESULT 3  
X40054  
ID X40054 standard; DNA; 852 BP.  
AC X40054:  
DT 02-JUL-1999 (first entry)  
DE Colon cancer associated gene.  
KW Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KW prostate cancer; ss.  
OS Homo sapiens.  
PN M09904265-A2.  
PD 28-JAN-1999.  
PE 15-JUL-1998; U14679.  
PR 22-JUN-1998; US-102322.  
PR 17-JUL-1997; US-896164.  
PR 10-OCT-1997; US-061599.  
PR 10-OCT-1997; US-061765.









PA (ARNOT) ARNOT D E.  
 PI Arnot DE, Enea V, Nussenzwei RS, Nussenzweig V;  
 DR WPI: 87-037250/05.  
 DR P-PSDB: P70708.  
 PT New Plasmodium vivax circumsporozoite protein - and synthetic  
 PT peptide(s) cont. its dominant epitope, useful in anti-malarial  
 PT vaccines  
 PS Disclosures; fig. 3; 32pp; English.  
 CC The circumsporozoite protein of P. vivax sporozoites encoded by  
 CC this gene is useful in the construction of a central domain of P70704  
 CC vaccine. The sequence consists of a central domain of P70704  
 CC repeated 19 times plus N- and C-terminal regions (practically  
 CC homologous with the corresponding domains of P. cynomolgi and P.  
 CC knowlesi. See also P70704-07, P70709 and N71065.  
 SQ Sequence 1908 BP; 677 A; 353 C; 470 G; 408 T;

Query Match 3.0%; Score 46; DB 1; Length 1908;  
 Best Local Similarity 44.8%; Pred. No. 0.0093;  
 Matches 261; Conservative 0; Mismatches 315; Indels 6; Gaps 2;

QY 601 caggtgaaaaggtgtacaaaagaatttcagctattatcaactcggaggaattcaa 660  
 DB 401 CAGAACCAAAATCCACGTAAATATAGCTGAACAACCGAGACAGACAGATGAC 460  
 QY 661 agccagctgtcttcgggaattggtcacaaggtgtgtgtggaatcgtatcaag 720  
 DB 461 AGCCAGCAGGAGAGACAGACAGATGACGCCAGAGGTATAGAGATGACACACAG 520  
 QY 721 ctgaataaacaacatggaga---aggtatgagcagaagaagagcagagttcaga 777  
 DB 521 CAGAGAGATAGAGCAGCTGACACACAGAGAGATAGAGATAGAGATGACAGACACA 580  
 QY 778 gcaaaaactaacagaggtcttgaataaagatcagaataatgagaattgagatgaa 837  
 DB 581 GAGCAGATGACACACAGAGAGAGACAGATGATGACCAACCGAGAGACAGAG 640  
 QY 838 ctggtatgaaagcaaacacacttggaaacagagcagaagagagcccttggcagagag 897  
 DB 641 CAGATGACCAACCGAGAGGTATAGAGAGCTGACCAACCGAGAGGTATAGACAGCTG 700  
 QY 898 gagtgcttgtaactaacagaactcgtgtggcgaatctgagcaacaacttgacctaccaga 957  
 DB 701 GACCAACCGAGAGATAGAGACAGATGACAGCCAGAGAGATAGAGATGACAGCTGACAGC 760  
 QY 958 tctgaatagctcaactcagcagcaagaaaaggtatcatatgataattggaagtta 1017  
 DB 761 CAGCAGAGATAGAGCAGATGACAGCCAGAGAGATAGAGCAGCTGAGACAGCAGCAG 820  
 QY 1018 cagaagaagaatgagaattggagaaacagtgtgtcagcactggagagatcagtgagagc 1077  
 DB 821 GAGATAGAGAGATGAGACAGCAGCAGAGATAGAGAGCTGAGAGCAGCAGCAGAGA---G 877  
 QY 1078 atgaagaagaaggtcagagcgtgtgataagcagcagccagccagcagcagctgtgtg 1137  
 DB 878 ATAGAGCAGCTGAGCAGCCAGCAGAGATAGAGCAGCTGAGCAGCCAGCAGAGATAGAG 937  
 QY 1138 cagctcctcagcagagcagaaccagcttctcctgtgagagagcag 1179  
 DB 938 CAGCTGAGACCGCAGCAGAGATGCTGAGCTGAGCAGCAGCAG 979

## RESULT 12

V21511  
 ID V21511 standard; DNA; 6755 BP.  
 AC V21511;  
 DT 17-AUG-1998 (first entry)  
 DE Staphylococcal bacteriocin BacRI operon.  
 KW BacRI; bacteriocin; antimicrobial; antibacterial; antibiotic;  
 KW Moraxella bovis; infectious bovine keratoconjunctivitis; cancer;  
 therapy; ds.  
 OS Staphylococcus aureus strain UT0007 (ATCC 55800).  
 PN M09812319-AL.

PD 26-MAR-1998.  
 PE 18-SEP-1997; U16758.  
 PE 17-SEP-1997; US-931999.  
 PR 19-SEP-1996; US-710561.  
 PA (UNIV ) UNIV KANSAS STATE RES FOUND.  
 PI Crupper SS, Iandolo JJ;  
 DR WPI: 98-230316/20.  
 PT Therapeutic proteolaceous substances from Staphylococcus aureus -  
 PT useful to inhibit growth of wide range of prokaryotic or eukaryotic  
 PT cells, e.g. Moraxella bovis causing infectious bovine  
 PT keratoconjunctivitis  
 PS Claim 2; Page 19-23; 38pp; English.  
 CC This polynucleotide comprises the bacteriocin BacRI operon of  
 CC Staphylococcus aureus UT0007. The sequence of the BacRI operon  
 CC was determined by N-terminal sequencing of purified BacRI peptide  
 CC (see M54171), with back-translation and plasmid analysis. The  
 CC BacRI operon includes the BacRI gene (see V21510), a homologue of  
 CC the cym gene of the cytolytic operon of Enterococcus faecalis  
 CC whose function is involved in the maturation of pre-cytolysin,  
 CC an ATP-transporter gene, bio1 and bio2 genes related to  
 CC lactococcal biosynthesis and modification, and a gene involved in  
 CC immunity function. BacRI peptides can be produced by construction  
 CC of an expression vector containing an oligonucleotide or operon  
 CC coding for BacRI, and use of the vector to transform host cells for  
 CC BacRI expression. The entire BacRI operon has been cloned into  
 CC plasmid pUB110, and Bacillus subtilis transformants secreted the  
 CC recombinant BacRI peptide into the medium. Bacteriocin BacRI is  
 CC active against many Gram-positive and Gram-negative organisms such  
 CC as Bordetella bronchiseptica, Pasteurella multocida and  
 CC Staphylococcus aureus; Moraxella bovis, causing infectious bovine  
 CC keratoconjunctivitis, is especially sensitive. BacRI can also be  
 CC used as an anti-cancer agent.  
 SQ Sequence 6755 BP; 3903 A; 1252 C; 1600 G; 0 U;

Query Match 2.9%; Score 44.8; DB 1; Length 6755;  
 Best Local Similarity 42.7%; Pred. No. 0.037;  
 Matches 286; Conservative 0; Mismatches 382; Indels 2; Gaps 1;

QY 461 gacttgaanaaagaacttgatctcagcaagaagaagcagcttgaaagaacatgtga 520  
 DB 3172 GAACGAAACCAAAACGCGCAACGCAACCAAAACCAACCGCA--GGGAAAGACAGACGA 3229  
 QY 521 aaaaagaataacgaagaagaggtacatggtgataaatgtgtatctgtctcaga 580  
 DB 3230 AAGGGGCAAGCAGAGGCAACGACAAAGAAACAAAGCAAAAGCCGCAACCAAAAGAG 3289  
 QY 581 atattcccaactggaagccaggttgaaaggttacaaagaagaattcagctatta 640  
 DB 3290 AGAAGCACCCACGAAAGCGCCAGCAGAAAGACGAAACGAGAAAGAAACCAAAACGAA 3349  
 QY 641 atcaactgaggaagaattcaagcagctgtctcgtggaatgagatgataaagtgat 700  
 DB 3350 GCCACGACCGAGAGAGCAAAAGAGAGACAGAAAGAGAAAGGGCGCAAAAGAGAGA 3409  
 QY 701 gtgagaagaatgcgtatcagctgataaacaacatcgagaagaagtgcagagaag 760  
 DB 3410 AAAAAAAGGAAAGAAACAAAGCGAAAGAAAGAGGGGAGAGAGAAAGAAAGAGAGACA 3469  
 QY 761 agcacagagagttcagaagcaaaaactaacagggatcttgaataatgaatcagaatag 820  
 DB 3470 AAAAAACCGAGAGAGAGAAAGAAAGGAAAGGAAAGAAAGGAAAGGAAAGGAAAG 3529  
 QY 821 agaatgagaataagaactgagatgaaagcaacaacacttggaaacggagcagagaag 880  
 DB 3530 ACAAGACGAG 3589  
 QY 881 cagccctgcagagagagagtgctgtgactaacagagactgctgggaggaatctgagacc 940  
 DB 3590 GAAAAAAGCAG 3649  
 QY 941 aactgacactcaccagatctgaaatagctcaactcagttcaagaagaaggtatatacatg 1000



```
Db 3650 AAAAAGCCGCAACCAACAGAGAAAAAACCAAAAGAGAAAGAAAAAAGAAAAAGCGAG 3709
Qy 1001 ataattgggaagtctacagagaagaatgaagaattggaggaacagtggtccgcatg 1060
Db 3710 CAAAAAGCGCAAAAAACGGAGAAACAAAGAAAAAACCAAAAAAGGAAAAAGCAAAA 3769
Qy 1061 ggaagtcacagagcagatgaagcaaaagctaaagcagctgataagcagcagcgcca 1120
Db 3770 AACGGGAGACAGCCACCAAAAAAAGGACAGAGAAAAAGCAAAAGAAAAACCAACAA 3829
Qy 1121 cagccagca 1130
Db 3830 AAGAGAAAGCA 3839

RESULT 13
ID T05868 standard; DNA; 3399 BP.
AC T05868:
DE 14-AUG-1996 (first entry)
KW Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.
KW Chicken leucocytozoan; Immunogen; recombinant vaccine; protection;
OS Chicken leucocytozoan.
FH Key Location/Qualifiers
FT cds 1 3399
FT /*tag= a
FT misc_feature 1150..3218
FT /*tag= b
FT /note= "fragment referred to in the claims, for
FT use as insert in a recombinant vaccine
FT against chicken leucocytozoan disease"
PN J07284392.A.
PD 31-OCT-1995.
PR 19-APR-1994; 080643.
PR 19-APR-1994; JP-080643.
PA (DOBU-) DOBUTSUJO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
PA (KITA) KITASATO KENKYUSHO SH.
DR WPI: 96-006311/01.
DR P-PSDB: R97866.
PT Chicken leucocytozoan immunogenic protein - used in a recombinant
PT vaccine against chicken leucocytozoan disease
PS Claim 6; Page 6-9; 35pp; Japanese.
CC T05868 encodes a chicken leucocytozoan immunogenic protein, this DNA
CC or a fragment of it can be used in a recombinant vaccine to immunise
CC against chicken leucocytozoan disease. The DNA is used in a vector
CC and operatively linked to an expression regulatory sequence as in
CC standard practice.
SQ Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T;
```

```
Db 2642 TAACACATGAAAGAGAGAAAAAGTAACACATGAAGAGAAAGAAAAAGTAATACATGAAG 2701
Qy 719 agctgaataaaaccacatggaagatgagcgagaaaggacacagaggttcagag 778
Db 2702 AAGAAAGAAAAAGAGAGATGAGAGAAAGAAAGAAAGAGAAAGAGAGAGAAAG 2761
Qy 779 caaacaactcaacggagctcttgaaatgaagatcaggaatagaagaattggaatagaac 838
Db 2762 AAGAAAGAGAGATGAGAGAGAGAGAAAGAAAGAAAGAAAGAAAGATGAGAGAGAAAG 2821
Qy 839 tggatgaagcaacaacacacttgaacaggaagcaggaagcagccctgagcagaagag 898
Db 2822 AAGAAAGAAAAATGAGAGAGAGAGAAAGAAAGAAAGAAAGAAAGAGAGAGAGAAAG 2881
Qy 899 agtccttgactaacagaactgctggcggaactctgaagcacaactgcacctcacagat 958
Db 2882 AAGACATGAAAGAGAGAGATGACATGAAAGAGAGAGAGAAAGAAAGTAACATGAAAGAG 2941
Qy 959 ctgaatagctcaactcagtcaggaagaaaggatatacatatgataattggaaagttac 1018
Db 2942 AAGAAAGAAAGTAACACATGAGAGAGAGAAAGTAACATATGAGAGAGAGAGAGAGAG 3001
Qy 1019 agagaagaatgaagaattggaagacagtggtccagcatggaagatgacatgagacga 1078
Db 3002 TAACACATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3061
Qy 1079 tgaagcaaaagcgaagcagctggaatgaagcagcagcagcagcagcagcagctgtgc 1138
Db 3062 AAGAAAGAAAGTAACATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3121
Qy 1139 agctcctcagcagagagacccgctctcctctggaagcagagcgtctcgggaaggtg 1198
Db 3122 AAAAAAGTAATGAGAGAGAGAGAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3181
Qy 1199 accggtcggagcccaagttaccacacatgcagcacaatctgat 1239
Db 3182 AAGGGGGGGGGCCATGTGTACCCGGATCCTCGAATTCGAT 3222

RESULT 14
ID T46187 standard; DNA; 3586 BP.
AC T46187:
DE 14-APR-1997 (first entry)
DE DNA encoding cell protein homologous to the SV40 large T antigen.
KW Cellular homologue of the SV40 large T antigen; CHLA-1;
KW natural killer lytic associated protein; tumour suppressor;
KW cancer; viral infection; therapy; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT exon 1..3069
FT /*tag= a
FT /codon.start= 48..50
FT intron 3070..3177
FT /*tag= b
FT exon 3178..3586
FT /*tag= c
FT /note= "in-frame stop codon at 3285-3287"
PN W09638540-A1.
PD 05-DEC-1996.
PR 30-APR-1996; U07691.
PR 30-MAY-1995; US-453701.
PA (RERE-) RES DEV FOUND.
PI Fung YK;
DR WPI: 97-034359/03.
DR P-PSDB: W01535.
PT DNA encoding cellular protein homologous with SV40 large T antigen -
PT useful for regulating initiation of DNA replication, e.g. for
PT treating tumors and viral infection
PT Example 10; Fig 11A; 59pp; English.
CC A DNA sequence (T46187) encodes a cellular homologue (W01535)
CC of the SV40 large T antigen (CHLA), characterised as a regulator of
CC initiation of DNA replication in cells that is useful e.g. in
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 19, 2000, 22:33:29 ; Search time 3256.04 Seconds  
(without alignments)  
1615.792 Million cell updates/sec

Title: US-09-502-945-3  
Perfect score: 1298  
Sequence: 1 ggcgcgcgaatgactgcga.....gaagaagaagaagaaaaa 1298

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 segs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.



QY 539 ctggaagcagaagaccccttgagcagaagaatttcaacccctccataagaagca 598  
||||| | ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 544 CTGGAAGGACAGACTCCCTTGAGCAGCANGAATTTTCACTTCTCATNGACAAGCA 603

QY 599 gccagtacagacccttactgacccctgtgaaaaagc 637  
| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 604 GNCAGTGACAGACCTTTACTGACCCCTGTGAAAGGCC 642

RESULT 2  
AA824593 507 bp mRNA EST 18-MAR-1998  
LOCUS 0c83d10.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1356307,  
DEFINITION mRNA sequence.  
ACCESSION AA824593.1 GI:2896615  
VERSION AA824593  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 507)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1406820.  
CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbfp/image/image.html](http://www.bio.llnl.gov/dbfp/image/image.html)

Insert Length: 2105 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 491.  
Location/Qualifiers  
1. 507  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1356307"  
/clone\_lib="NCI\_CGAP\_GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH108"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker. Site1: Not I; Site2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD+),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
[5'-TGTTACCAATCTGAATGGAGCGGCCCTCATTTTCTTTTCTTTT-  
3']. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 90 a 120 c 117 g 180 t  
ORIGIN

Query Match 32.0%; Score 415; DB 39; Length 507;  
Best Local Similarity 91.3%; Pred. No. 5, 4e-68;  
Matches 461; Conservative 0; Mismatches 42; Indels 2; Gaps 2;

QY 363 tactctgaacaaagaagagattgaacggaatccacagaga--tagcatcaataaaag-- 421  
||||| | ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 507 TACCTCTTAACAAGAAAAATTGAACAGATCCACAGAGAGTAGATTCAGTAACCT 448

QY 421 cacaagctctctcaaatggaacctgctgctcctgaagaacccgagcagcagctg 480  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 447 CACAGGCTCTCTCCAATGGAGCCCTATCATCTGAAAGAACAGCAGCAGCAGCTGG 388

QY 481 ttttcccttgagaagaagagagcagccatgctcccatgaacatgctcagtgc 540  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 387 TTTTTCACCTGGGGAGAGACACCCACCCATTCCTCCATTTGAACATGCGCTCAGTGGCT 328

QY 541 ggaagcagaagaccccttgagcaggaatttcaacccctccatgaagaacagc 600  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 327 GGAAGGCAAGACTCCCTTGAGCAGGAAATTTTAACTCTTCATTAAGCAAGCAGC 268

QY 601 cagtgcagacccttactgacccctgtgaaaaagcctctccgcagccatgagctag 660  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 267 CAGTGAAGATCCTTTACTGACATCCCATGGAAGAGCGCTCTCCAAAGCATGAGCCTGG 208

QY 661 aagaagcagaatgacgacgagcagagctcagaggtcggtcgctcagctactatg 720  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 207 AAGAGCAAAAGATGACACCGAGAGGCTTCAGAGGCTCGGCTCTGACGTCTACTATG 148

QY 721 aagcagaagctgagaagaagaataatnaagttaaagtatcaaaagctgaga 780  
| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 147 AGGCCAAGGCTGAAAGAGAGAAATCAAAAGTAAAGTATCAAAAGTGTGAAGA 88

QY 781 aagaagcagaagaagccctaaagaagcttgaagcgtcggaaggttaaccagct 840  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 87 AAGGAAGGCGCAAGAAAGCCTTAAAGAGCTTGAAGCAGCTACAGAAAGTTATCAACTG 28

QY 841 cgcactagaagaagaagaaga 865  
||| ||||| |||||  
Db 27 TGGCATTGGAAGAAATGAAAAAAA 3

RESULT 3  
LOCUS AA151187 392 bp mRNA EST 10-DEC-1996  
DEFINITION z003c11.r1 Strataene colon (#937204) Homo sapiens cDNA clone  
IMAGE:566612 5', mRNA sequence.  
ACCESSION AA151187  
VERSION AA151187.1 GI:1719307  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 392)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,  
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rickin, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,  
and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
TITLE JOURNAL MEDLINE  
ON Sep 12, 1996 this sequence version replaced gi:1398021.  
CONTACT: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [estewatson.wustl.edu](mailto:estewatson.wustl.edu)  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -28m13 rev2 from Amersham.  
Location/Qualifiers  
1. 392  
/organism="Homo sapiens"



```
RESULT 5
LOCUS T16360/c 361 bp mRNA EST 25-JUL-1996
DEFINITION NM1193 Normalized Infant Brain, Bento Soares Homo sapiens CDNA
ACCESSION T16360
VERSION T16360
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 361)
AUTHORS Berry,R., Stevens,T.J., Walter,N.A.R., Wilcox,A.S., Rubano,T.,
Hopkins,J.A., Weber,J.G., Goold,R., Soares,M.B. and Sikela,J.M.
Gene-based Sequence Tagged Sites (STSs) as the basis for a human
gene map
JOURNAL Nature Genet. 10, 415-423 (1995)
MEDLINE 95400322
COMMENT Contact: Sikela JM
Department of Pharmacology
University of Colorado Health Sciences Center
Box C236, 4200 E. 9th Ave, Denver CO 80262-0236
Tel: 3032708637
Fax: 3032707097
Email: nikki@tally.uchsc.edu
Primers: CCTTCTTCACGACTTGTG , TGACAGATCCTTTACTGACTCC
Seq primer: -21M13 universal.
FEATURES
source
1..361
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="879f05:13:13q14.1-13q22.3"
/clone_lib="Normalized Infant Brain, Bento Soares"
/lab_host="E. coli DH5-alpha"
/note="Vector: pA, M13-derived; Site_1: HindIII; Site_2:
NotI; The normalized infant brain library, constructed by
Bento Soares, Columbia University, was oligo-(dT) primed
and directionally cloned into an M13-derived plasmid using
total brain mRNA from a 72-day old human female afflicted
with spinal muscular atrophy. The library was normalized
as described elsewhere."
BASE COUNT 63 a 92 c 83 g 119 t 4 others
ORIGIN
Query Match 24.6%; Score 319; DB 20; Length 361;
Best Local Similarity 92.7%; Pred. No. 3.7e-50;
Matches 331; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
```

```
RESULT 6
LOCUS A0028402 388 bp DNA GSS 30-JUN-1998
DEFINITION CIT-HSP-2313E15.TF CIT-HSP Homo sapiens genomic clone 2313E15,
genomic survey sequence.
ACCESSION A0028402
VERSION A0028402
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 388)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,M.K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2313E15.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
FEATURES
source
1..388
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CIT-HSP"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT 122 a 81 c 89 g 96 t
ORIGIN
Query Match 24.3%; Score 315.2; DB 94; Length 388;
Best Local Similarity 91.5%; Pred. No. 1.9e-49;
Matches 354; Conservative 0; Mismatches 31; Indels 2; Gaps 2;
```

[illegible]

Db	Accession	Version	Keywords	Source	Organism	Journal	Comment
Db	386	CCCTGGGGAGAGACAGCCACCATTGCTCCCATTAAGATCGGCTCAGTGGCTG	327				
Qy	543	aagcaagaact-cccctggagcaggaatttccaacctccatcaaaagcaagcagc	601				
Db	326	AAGCAGAAACTCCCTGTGAGCAGGAATTTTAACTCTCCATTAAGAACAAACAGCC	267				
Qy	602	atgacagacccttactgacccctgttgaaaggcctctctccgagccatgagcctaga	661				
Db	266	AGTGACAGATCCCTTACATGACCTCCATGGAAGGAGGCTCTCTCAAGGCATGAGCTGGA	207				
Qy	662	agaagcaaatgctgcagcagcagcagccttcaagaggctcgagctgctgagctactatga	721				
Db	206	AAGGCAAAAGATGCACCGACGACGAGCTTCAGAGGCTCGGCTCTGCACTCTACTATGA	147				
Qy	722	ngcgaagctcgaagagaagaanaatcnaaagtaaaagtaacaaagtcgtgaagaa	781				
Db	146	GGCCAAAGCTCGAAGAAAGAGAAAGAAATCAAAAGTAAAGATCAAAAGTGTGAAGAA	87				
Qy	782	aggaagaagcgaagaagaagccctaaagaagttgagcagctgcggaagtgtaacagctgc	841				
Db	86	AGGAAAGGCGCAAGAAAGCCTTAAAGAGTTTGAGCAGCTACAGAAAGTAAATCCAACTGT	27				
Qy	842	cgactatagaagaagaagaagaaga	865				
Db	26	GGCATTGGAGAAATGGAATAAA	3				
RESULT	8	AA954862	327 bp	mRNA	EST	07-MAY-1998	
LOCUS	AA954862	op20c03.s1	NCI-CGAP.Co12	Homo sapiens	cdna clone IMAGE:1576228	3'	
DEFINITION		similar to TR:Q92555	Q92555	MYELOBLAST	KIAA0266	;	mRNA sequence.
ACCESSION		AA954862					
VERSION		AA954862.1	GI:3118557				
KEYWORDS		EST.					
SOURCE		human.					
ORGANISM		Homo sapiens					
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;					
		Eutheria; Primates; Catarrhini; Homiidae; Homo.					
REFERENCE		NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.				
AUTHORS		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),					
TITLE		Tumor Gene Index					
JOURNAL		Unpublished (1997)					
COMMENT		On Jan 19, 1998 this sequence version replaced gi:2151515.					
		Contact: Robert Strausberg, Ph.D.					
		Tel: (301) 496-1550					
		Email: Robert.Strausberg@nih.gov					
		Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.					
		Emmert-Buck, M.D., Ph.D.					
		CDNA Library Preparation: Stratagene, Inc.					
		DNA Sequencing by: Washington University Genome Sequencing Center					
		Clone distribution: NCI-CGAP clone distribution information can be					
		found through the I.M.A.G.E. Consortium/MLN, at:					
		www-bio.lnl.gov/dbp/image/image.html					
		Seq primer: -40m13 fwd. ET from Amersham					
		High quality sequence stop: 311.					
FEATURES		Location/Qualifiers					
source		1..327					
		/organism="Homo sapiens"					
		/db_xref="taxon:9606"					
		/clone="IMAGE:1576228"					
		/clone_lib="NCI-CGAP.Co12"					
		/sex="mixed"					
		/tissue_type="colon tumor"					
		/lab_host="SOLR (kanamycin resistant)"					
		/note="Organ: colon; Vector: Bluescript SK-; Site: 1;					
		Ecotri, Site_2: XhoI; Cloned unidirectionally. Primer:					
		0150 df. Pooled colon tumors. 5' adaptor sequence: 5'					
		GAAATGGGACGAG 3' 3' adaptor sequence: 5'					





Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,  
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,  
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,  
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,  
Fraser,C.M., and Venter,J.C.  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
Nucleic Acids Res 23 (1995) 377-384 (1995)

JOURNAL  
MEDLINE  
COMMENT  
12140200  
On May 9, 1995 this sequence version replaced gi:803089.  
Other ESTs: THC192511  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
7112 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. 268  
/organism="Homo sapiens"  
/db\_xref="ATCC (lnhost):177963"  
/db\_xref="taxon:9606"  
/clone\_lib="HSCI72 cells 1"  
/cell\_type="fibroblast"  
/cell\_line="HSCI72 (16PDL)"  
/note="stage=fetal"  
/note="organ: lung; Vector: pBluescript SK-; Site\_1:  
EcoRI; Site\_2: XhoI"

BASE COUNT 82 a 45 c 84 g 56 t 1 others  
ORIGIN

Query Match 20.3%; Score 264; DB 32; Length 268;  
Best Local Similarity 98.5%; Pred. No. 6e-40; Indels 0; Gaps 0;  
Matches 264; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 gctcaaatgactgcgaacccgcttcgagagagccttcgtcttgagccanagagaagaa 64  
Db 1 GCTCAAAATGACTGCGAACCGCTTCGAGAGAGCCTTCGCTTGAGCCACAGAGAA 60  
QY 65 ctacggatttcgcaaaagactacctcttgagtgagagtgagaatgaggggacatgat 124  
Db 61 CTACGGATTTCGCAAAAGACTACCTCTTGAGTGAAGTGAAGTGAAGGCGACATGAT 120  
QY 125 gggagagagaagactcnaaagcttcggaagcaatcagctcccttgatggaagaatag 184  
Db 121 GGAGAGAGAAAGCAATCAAAAGCTTCGAAACCAATCAATGATGAGAAAGATAGG 180  
QY 185 cggaaatgctgctgaagagctcgaagctcgaagtgatgagagtgatcgaatgctct 244  
Db 181 CGGAATTTGCTGAGAGCTCTGAGGCTAGTCTGAAGGTCTCAAGTCAATGCTCACT 240  
QY 245 gaagatcaggaagaagctgctctg 272  
Db 241 GAAGATCAGAGAGAAAGCTGCTCTG 268

RESULT 11  
AW291601/c 330 bp mRNA EST 16-JAN-2000  
LOCUS  
DEFINITION  
IMAGE:2724730 3', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
human.

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT  
1 (bases 1 to 330)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jul 9, 1999 this sequence version replaced gi:5434385.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. cDNA library preparation: M.B. Soares Lab Clone distribution:  
NCI-CGAP clone distribution information can be found through the  
I.M.A.G.E. Consortium/LNLN at:  
[www.bio.lnl.gov/dbir/image/image.html](http://www.bio.lnl.gov/dbir/image/image.html)  
Seq primer: M13 Forward  
POLYA=yes.  
Location/Qualifiers  
1. 330  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NCI CGAP Sub4"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pTZ19-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The  
NCI-CGAP Sub4 library is a subtracted library derived from  
the NCI-CGAP Sub2 library which is a subtracted library  
derived from the NCI-CGAP Sub1 library, which is a  
subtracted library derived from BI. BI constitutes a  
mixture of 21 normalized or subtracted NCI-CGAP  
libraries: NCI-CGAP-Co4, NCI-CGAP-Pr22, NCI-CGAP-Pr28,  
NCI-CGAP-Co10, NCI-CGAP-Co16, NCI-CGAP-Kid5,  
NCI-CGAP-Kid12, NCI-CGAP-Kid3, NCI-CGAP-Co8, NCI-CGAP-CLL1,  
NCI-CGAP-Lei12, NCI-CGAP-Brn23, NCI-CGAP-Lu5,  
NCI-CGAP-Lu24, NCI-CGAP-Lu19, NCI-CGAP-GC4, NCI-CGAP-GC6,  
NCI-CGAP-Brn25. These 21 libraries were pooled and a  
single-stranded DNA preparation of the resulting mixture  
was used as a tracer in a subtractive hybridization with  
a diver whose composition is detailed below:  
NCI-CGAP Kid3 pool 1: LLM 3334-3337, 3682-3683,  
3798-3803 (IMAGE Clonoids 1322376-1323911,  
1456008-1456775, 1500552-1502855) NCI-CGAP Kid5 pool 1:  
LLM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids  
1323912-1325831, 1471368-1472903, 1492104-1493255)  
NCI-CGAP Lu5 pool 1: LLM 3575-3582, 3851-3854 (IMAGE  
Clonoids 1414920-1417981, 1520904-1522439) NCI-CGAP GC4  
pool 1: LLM 3164-3167, 3716-3720, 3733-3735 (IMAGE  
Clonoids 1257096-1258631, 1469064-1470983,  
1475592-1476743) NCI-CGAP Pr22 pool 1: LLM 2457-2459,  
2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759,  
1101192-1101959, 1217928-1220615) NCI-CGAP Co10 pool 1:  
LLM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255,  
114584-1145351) Subtraction was performed as previously  
described (Bonaldi, Lennon & Soares (1996): Normalization  
and Subtraction: Two Approaches to Facilitate Gene  
Discovery. Genome Research 6, 791-806.)  
TAG-NCI CGAP\_Brn23  
TAG-TISSUE-drain  
TAG\_SEQ-ATATC"

BASE COUNT 59 a 78 c 89 g 104 t  
ORIGIN

Query Match 20.3%; Score 264; DB 79; Length 330;  
Best Local Similarity 87.5%; Pred. No. 6.2e-40;  
Matches 288; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY	419	CGCAGAGCTCCTCCCAATGAGGCGCTATCGCTCTGAAGAACCGGAGGCAAGAGCT	478
Db	330	CTCACAGCTCCTCTCCCAATGGAGCCTATCATCTCTGAGAACGACGACGACGACT	2711
QY	479	ggttttccctctgagaagaagagagccatctgtcccatlgaacatgltgctcaatg	538
Db	270	GGTTTTTCCCTCGGGGAAGGAGCGACGACCATGCTCTCCATTGACATCGCTCACTGG	2111
QY	539	ctggaagcaagaacatccccctggagcagagaaatttcaactctctcgaabaagaacagca	598
Db	210	CTGGAAGGCAGAACTCCCTCGAGCGACGGAATTTTAACTCTCTCATGAACAAAGCA	151
QY	599	gcccagtgacagacccttacttgacccctgtggaaagcctctctccgagccatlgagct	658
Db	150	GCCAGTGCACATATCTTTACTAGCATCCCATGGAAGGCGCTCTCTCCAAACCATGAGACT	91
QY	659	agaagagcaagaatgtcagcagcagagagcttcagaagcctcgagctctgtaactacta	718
Db	90	GGAAGAGGCAAAAGTATGCACCGAGAGAGCTTCAGAGGAAGAGGCTCTACGATCTAACTG	31
QY	719	tgangccaagctctcgaagagagagaagaaa	747
Db	30	TGACACCATGCAGAAAAAAAAAAAAA 2	

RESULT	12
R42009/c	
LOCUS	R42009 449 bp mRNA
DEFINITION	Y905004.s1 Soares infant brain INIB Homo sapiens cDNA clone
IMAGE:	31336 3', mRNA sequence.

ACCESSION	R42009
VERSION	R42009.1
KEYWORDS	EST.
SOURCE	human.

REFERENCE  
AUTHORS

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia:  
Eutheria: Primates: Catarrhini: Homiidae: Homo.  
1 (bases 1 to 449)  
Hiller, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Merri, M.,  
Parsons, J., Rifkin, L., Ronfling, T., Soares, M., Tan, F.,  
Trevisansk, E., Waterston, R., Williamson, A., Wohlmann, P. and  
Wilson, R.

TITLE	The WashU-Merck EST Project
JOURNAL	Unpublished (1995)
COMMENT	On May 8, 1995 this sequence version replaced gi:800233.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
Insert Size: 1364  
High quality sequence stops: 322 Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert length: 1364 Std Error: 0.00  
Seq primer: Promega -21ml3  
High quality sequence stop: 322.

**FEATURES**  
**SOURCE**

```

/organism="Homo sapiens"
/db_xref="GDB:403683
/db_xref="taxon:9606"
/clone="IMAGE:31336"
/clone_11b="Soares Infant brain IN1B"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain, Vector: lambda
Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer 15',
LACGTGGAGAAATTCGCGCGCGGACGAATTTTTTTTTTTTTTTT 3'];

```

double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the  $\lambda$ gt10 BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fátima Bonaldo."

Query Match	17.1%;	Score 222.6;	DB 22;	Length 449;
Best Local Similarity	85.1%;	Pred. No. 3.3e-32;		
Matches 343;	Conservative 0;	Mismatches 46;	Indels 14;	Gaps 8

QY 477 ctggttttccctggaagagagcc--agccattgctcccatgacatg--tgc 531

Db 405 CTGGTTTTCCTGGGGAAGGAGCAGCCCACTTGCTCCCATTTGAACATGCCGCC 346

QY 532 tcagtgcgtgaag-cagaactccctgag-caggaatttcaactct--ccat 587

D5 345 TCAGTGGCTGGAGGCGCAAGACTCCCTGGAGCCAGGAATTTTAACCTCCCTCCCAT 286

588 aagaacaagcag-ccagtgcacaga-cccttactgacccctgtgaaaagcctctcc 645

DB 285 AAGAACAGCAGCCAGTGGACAGATCCCTTTACTGACTCCCATGGAAAGGCTCTCTCC 220

040 yagccagag--ccagaaagcagagcagagcagagagcagag /02

[illegible][illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

g  
h  
i  
j  
k  
l  
m  
n  
o  
p  
q  
r  
s  
t  
u  
v  
w  
x  
y  
z

## RESULT 13

LOCUS	AA100660	521 bp	mRNA	EST	31-JUL-1997
LOCUS	AA100660	521 bp	mRNA	EST	31-JUL-1997

IMAGE:511856 5', mRNA sequence.

VERSION	AA100660.1	GI:1646978
KEYWORDS	EST	

**SOURCE**  
**ORGANISM**

Eutheria: Primates: Catarrhini: Hominoidea: Homo  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

## REFERENCES

Chisose, S., Dietrich, N., Dubaque, T., Favello, A., Gish, M., Hawkins, M., Hultman, M., Kucab, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M. B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)

TITLE  
JOURNAL  
MEDLINE  
COMMENT

97044478  
Contact: Wilson RK

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [estewartson.wustl.edu](mailto:estewartson.wustl.edu)  
This clone is available royalty-free through LINT ; contact the  
IMC Consortium ([info@imc.llnl.gov](mailto:info@imc.llnl.gov)) for further information.

Insert Length: 4133 Std Error: 0.00  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 404.

## FEATURES

## SOURCE

1. 521  
/organism="Homo sapiens"  
/db\_xref="GDB:3844351"  
/db\_xref="taxon:9606"  
/clone="IMAGE:511856"  
/clone.lib="Stratagene colon (#937204)"  
/tissue\_type="tumor"  
/cell\_line="T84 carcinoma cell line"  
/note="Organ: colon; Vector: pBluescript SK-; Site: 1:  
EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:  
Oligo df: T-84 colonic epithelial cell line. Average  
insert size: 1.0 kb. Uni-ZAP XR Vector. -5' adaptor  
sequence: 3' GAATTCGACGAG 3' -3' adaptor sequence: 5'  
CTCAGCTTTTTTTTTTTTTTTT 3' "

BASE COUNT 183 a 97 c 151 g 86 t 4 others  
ORIGIN

Query Match 16.5%; Score 214.8; DB: 28; Length 521;  
Best Local Similarity 87.2%; Pred. No. 9,7e-31;  
Matches 245; Conservative 0; Mismatches 35; Indels 1; Gaps 1;

Qy 630 gaaagagctctccgcagccatgagcttaagaaggaagatgagcagcagctt 689  
|||||  
Db 1 GAAAGGCGCTCTCCGAGCCATGAGCTGAAGGCAAGATCGACGA-MAGACTT 59  
Qy 690 cagaagagctcagctcagctcagctcagctcagctcagctcagctcagctc 749  
|||||  
Db 60 CAGAGGCGCTCGGCTCTCAGTCTCAGTATGAGCGCAAGGCTCGAAGAGAAATC 119  
Qy 750 naagttaaagatcacaaagtcgtgagaagaaggaagcgaagccctaaagag 809  
|||||  
Db 120 AAAAGTAAAGATATCACAAAGTCGTGAAGAAAGGCAAGAAAGCCCTAAAGAG 179  
Qy 810 tttaagagctcgaggaaggttaactcagctcagctcagctcagctcagctcagctc 869  
|||||  
Db 180 TTTTGAAGCGCTCGGAGGTTAATCCAGCTCAGCTGAGAGAACTGCAAAATTTGA 239  
Qy 870 agaagagagagagaagaagaacacaagaagaagaaga 910  
|||||  
Db 240 AAGGCAGATGATGAGAAAGATAGCCTTAAGCACCAAA 280

RESULT 14  
CNS005TE/c GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #  
DEFINITION BACRI2K22 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION AL060767.1 GI:4943573  
VERSION AL060767  
KEYWORDS GSS.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 997)  
Genoscope.  
TITLE Direct Submission  
AUTHORS Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
COMMENT Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osogawa and  
Aaron Mammeter in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

## SOURCE

1. 997  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/clone.lib="RPCI-98"  
/clone="BACRI2K22"  
/note="end : TET3"

BASE COUNT 89 a 99 c 13 g 258 t 538 others  
ORIGIN

Query Match 16.0%; Score 208.2; DB: 82; Length 997;  
Best Local Similarity 25.4%; Pred. No. 1.8e-29;  
Matches 132; Conservative 265; Mismatches 122; Indels 1; Gaps 1;

Qy 776 gaagaagaagagcgaagaaagccctaaagattgagcagctcgaggttaatcc 835  
|||||  
Db 983 RARCARRRRRRCARCARARARARARARARARARARARARARARARARAR 924  
Qy 836 agctcgactagaagaagaagaagaagaagaagaagaagaagaagaagaaga 895  
|||||  
Db 923 RRRGRRRRRGRRRRSAGCGRRGRRAAARARARARARARARARARARAR 864  
Qy 896 caagagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 955  
|||||  
Db 863 ARRR 804  
Qy 956 agaagagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1015  
|||||  
Db 803 ARARR 744  
Qy 1016 gaagagaagaagaataagaagaagaagaagaagaagaagaagaagaaga 1075  
|||||  
Db 743 RRARRRGAARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 684  
Qy 1076 gaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1135  
|||||  
Db 683 RRRGRRR 624  
Qy 1136 taaggaacnaagaagaagaagaagaagaagaagaagaagaagaagaaga 1195  
|||||  
Db 623 RRGARCGRR 565  
Qy 1196 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1255  
|||||  
Db 564 ARGAARR 505  
Qy 1256 agaaagaagaatataagaagaagaagaagaagaagaagaagaaga 1295  
|||||  
Db 504 GRRGRR 465

RESULT 15  
AV118269  
LOCUS AV118269 273 bp mRNA EST 30-JUN-1999  
DEFINITION AV118269 Mus musculus C57BL/6J 10-day embryo Mus musculus CDNA  
clone 2610207M07, mRNA sequence.

ACCESSION AV118269  
VERSION AV118269.1 GI:5300420  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 00:30:22 ; Search time 158.18 Seconds

(without alignments)  
1041.916 Million cell updates/sec

Title: US-09-502-945-3

Perfect score: 1298  
Sequence: 1 ggcgtgcgaatgactgcga.....gaagaagaagaagaanaa 1298

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 226296 seqs, 63486255 residues

Total number of hits satisfying chosen parameters: 452592

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/PCrus.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	215.4	16.6	7218	1 US-08-232-463-14	Sequence 14, Appl
2	95.4	7.3	19124	4 US-08-487-826B-13	Sequence 13, Appl
3	93.6	7.2	240	1 US-08-628-417-6	Sequence 6, Appl
4	91.6	7.1	2447	4 US-09-014-969-14	Sequence 14, Appl
5	86	6.6	291	1 US-07-922-723A-7	Sequence 7, Appl
6	86	6.6	291	1 US-07-799-828C-7	Sequence 7, Appl
7	86	6.6	291	1 US-08-074-215-7	Sequence 7, Appl
8	86	6.6	291	1 US-08-480-366-7	Sequence 7, Appl
9	86	6.6	291	3 US-07-952-277A-7	Sequence 7, Appl
10	85.8	6.6	2223	3 US-08-257-073-4	Sequence 4, Appl
11	84	6.5	494	3 US-08-332-766A-22	Sequence 22, Appl
12	82.8	6.4	376	3 US-08-623-906A-18	Sequence 18, Appl
13	82.8	6.4	43795	5 US-08-742-185-101	Sequence 10, Appl
14	82.4	6.3	4766	6 PCT-US93-07261-10	Sequence 10, Appl
15	82.2	6.3	1859	5 US-08-691-563C-46	Sequence 46, Appl
16	81.8	6.3	446	3 US-08-332-766A-26	Sequence 26, Appl
17	80.4	6.2	434	3 US-08-623-906A-6	Sequence 6, Appl
18	78.6	6.1	3211	4 US-08-574-959A-8	Sequence 8, Appl
19	78.6	6.1	3901	4 US-08-574-959A-6	Sequence 6, Appl
C 20	78.4	6.0	4698	1 US-07-807-043B-5	Sequence 5, Appl
C 21	78.4	6.0	4698	1 US-08-299-849B-5	Sequence 5, Appl
C 22	78.4	6.0	4698	4 US-08-142-368A-5	Sequence 5, Appl
C 23	78.4	6.0	4698	5 US-08-967-727-5	Sequence 5, Appl
C 24	78.2	6.0	285	3 US-08-630-822A-85	Sequence 85, Appl
25	78.2	6.0	285	4 US-09-005-069-85	Sequence 85, Appl
26	78	6.0	1236	3 US-08-741-134-5	Sequence 5, Appl
C 27	77	5.9	1598	5 US-08-785-420-3	Sequence 3, Appl

28	76.2	5.9	6243	4 US-09-056-075-1	Sequence 1, Appl
C 29	76	5.9	56516	4 US-08-996-306-1	Sequence 1, Appl
30	74	5.7	1434	4 US-08-903-801-2	Sequence 2, Appl
31	73.6	5.7	370	3 US-08-332-766A-8	Sequence 8, Appl
32	73.6	5.7	3337	1 US-08-072-610-1	Sequence 1, Appl
33	73.6	5.7	3337	3 US-08-719-822B-1	Sequence 1, Appl
34	73	5.6	1621	4 US-08-933-750C-96	Sequence 96, Appl
35	73	5.6	5394	5 US-08-688-376-1	Sequence 1, Appl
36	72.4	5.6	1866	5 US-09-173-581-13	Sequence 13, Appl
37	69.4	5.3	2261	1 US-08-272-882D-1	Sequence 1, Appl
38	68.6	5.3	1192	2 US-08-340-539A-3	Sequence 3, Appl
39	68.6	5.3	1192	3 US-08-461-592B-3	Sequence 3, Appl
40	68.6	5.3	1377	4 US-08-810-572A-1	Sequence 1, Appl
C 41	67.6	5.2	217	3 US-08-332-766A-4	Sequence 4, Appl
C 42	67.2	5.2	421	2 US-08-480-784-24	Sequence 24, Appl
C 43	67.2	5.2	421	2 US-08-483-553-24	Sequence 24, Appl
C 44	67.2	5.2	421	2 US-08-487-002-24	Sequence 24, Appl
C 45	67.2	5.2	421	2 US-08-483-554B-24	Sequence 24, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHNEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-Fls  
; US-08-232-463-14















```

PCT-US93-07261-10      RESULT 14
; Sequence 10, Application PC/TUS9307261
; GENERAL INFORMATION:
; TITLE OF INVENTION: PFEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John H. C. Blasdale
; STREET: One Giralda Farms
; CITY: Madison
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07940-1000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07261
; FILING DATE: 19930805
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,531
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Blasdale, John H. C.
; REGISTRATION NUMBER: 31,895
; REFERENCE/DOCKET NUMBER: DX0288K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-822-7398
; TELEFAX: 201-822-7039
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4766 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; STRAIN: Malayan Camp
; IMMEDIATE SOURCE:
; CLONE: p2b1.p12-1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..4766
PCT-US93-07261-10

```

	Query Match	6.3%	Score 82.4;	DB 6;	Length 4766;	
	Best Local Similarity	52.0%;	Pred. No. 3.4e-09;			
	Matches	229;	Conservative	0;	Mismatches 200;	Indels 3; Gaps 2.
OY	857	aagaaagaaggaagaaggaaggaagaagaagaacaaggagaagaagaaga	                   			916
Db	8	AAGAAAGGCGCTTTAAAACACAAAACTCAAAAAAAAAATGAAAAAGCAGAAATGCTTTAA	                   			67
OY	917	aggggaagggaagaaaaagaagaaggaagaaaagaagaagaa--agaagaagyg	                   			975
Db	68	GCAAAAGAATTAAAAGAACAAAAAAGACTACCCGCAGAAAGCAAAGATTTAAACA	                   			127
OY	976	aaggaagaagaagaagaagaagaagaactaagaagaagaagaagaataaga	                   			1035
Db	128	AAAGGAAGGCCAAGACTCTTCAGTGAATAAACTACTAAAGGAAAAAGTAAAGGGAAMC	                   			187
OY	1036	agggaagaagaagaagaagaagttnaagaagaagaagaagaagaagaagaaga	                   			1095
Db	188	ATTAAAGCAAAAAGAAATTAAGAAACATTTAAAAAAAAGAGCTTAGAAATCACAGAGA	                   			247
OY	1096	gaactnagaagaagaagaagaagaagaagaagaatagaacmagaagaaga	                   			1155
Db	248	AAAGAGCAAAAAATTAATAATTAAGATATATATCATGTAGAACATTAA--AGATTTAAGGA	                   			305

[illegible]

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

	Query Match	Similarity	52.7%	Score 82.2	DB 5	Length 1859
	Best Local	Similarity	52.7%	Pred. No. 3e-09		
	Matches	174	Conservative	0	Mismatches 155	Indels 0
QY	869	aagaaagagagagagaaagaaagaaacaaaggaagaagaagaagaagag				928
Db	184	AAAAGGAAAGAGGAGAGAGACACATTAAGGCGTACGAGCCAAAGGAACCTTAGCG				243
QY	939	aagaaagagagagagagaaagaaagaaagagagaaagaaagagagaaagaa				988
Db	244	AAAGCAAGAGAGGAAAGAGACAGAAAGTCACGAGAGAGAGAGGAAAGACACAGACACAA				303



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 00:41:50 ; Search time 243.1 Seconds  
(without alignments)  
2225.073 Million cell updates/sec

Title: US-09-502-945-5

Sequence: 1 cctggccgcgtgcgcgtcgc.....tatttccagcttaaaaaa 2162

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N.Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2162	100.0	2162	1	X40058	Colon cancer assoc
2	2078	96.1	2236	1	X39644	Renal cancer assoc
3	2078	96.1	2236	1	X40057	Colon cancer assoc
4	2025	93.7	2289	1	X40061	Colon cancer assoc
5	1905	88.1	2409	1	X40062	Colon cancer assoc
6	1148	53.1	1306	1	X40060	Colon cancer assoc
7	58.8	2.7	32207	1	V73805	SHV long unique c
8	58.8	2.7	137507	1	V19941	KSHV LTR DNA (nuc1
9	57.2	2.6	799	1	V55831	Nucleotide sequenc
10	57.2	2.6	9600	1	V21683	Vector plasmid pcm
11	57.2	2.6	10596	1	Q51731	Plasmid pCLISEBON f
12	57.2	2.6	10596	1	T40348	Plasmid pCLISEBON f
13	57.2	2.6	10596	1	X15650	Nucleotide sequenc
14	57	2.6	795	1	V55830	FLGA insert stablil
15	52	2.4	543	1	O23092	Antigen tc-7a gene
16	49.4	2.3	2744	1	O98470	Mispl-containing p
17	48.8	2.3	29392	1	V15422	Mouse poly ig rece
18	47.6	2.2	4000	1	T91902	Manose-1-phosphat
19	46.8	2.2	2338	1	Q14183	N.clavipes draglin
20	46.8	2.2	2338	1	V23249	Neplila clavipes s
21	46.4	2.1	2000	1	N71065	Gene encoding plas
22	46	2.1	243	1	O37279	Clone 2-9 used in
23	46	2.1	243	1	O95004	SSP 7.7.7.7.7.8.
24	46	2.1	243	1	V35842	Synthetic storage
25	46	2.1	243	1	V95939	Synthetic lysine-r
26	46	2.1	1390	1	V84524	Human secreted pro
27	45.6	2.1	12001	1	O76213	HSV L/ST region. H
28	45.6	2.1	234	1	O84832	Spinocebellar at
29	45.2	2.1	1505	1	O55750	Genomic clone G11F
30	45.2	2.1	2301	1	V20445	Human c-trk oncoge
31	44.6	2.1	117213	1	V62176	HSV-2 strain SB5 C
32	44	2.0	1686	1	O87587	DNA encoding Leuco
33	44	2.0	3512	1	X37085	FMF associated pro
34	43.6	2.0	154	1	O84835	Spinocebellar at

35	43.6	2.0	203	1	V17226	SCA2 gene CAG repe
36	43.6	2.0	203	1	V30271	Glutamine rich reg
37	43.6	2.0	913	1	T30271	Cotton fibre-spect
38	43.6	2.0	913	1	T13034	Cotton fibre-spect
39	43.6	2.0	913	1	T30253	Cotton fibre-speci
40	43.6	2.0	913	1	T62610	Cotton fibre speci
41	43.6	2.0	913	1	T70041	Cotton fibre-speci
42	43.6	2.0	1984	1	T13030	Cotton fibre-speci
43	43.6	2.0	1985	1	T30250	Cotton fibre-speci
44	43.6	2.0	1985	1	T70036	Cotton H6 gene and
45	43.2	2.0	756	1	X39869	Gastric cancer ass

## ALIGNMENTS

RESULT 1  
X40058  
X40058 standard; DNA; 2162 BP.

AC	X40058;	02-JUL-1999 (first entry)
DE	Colon cancer associated gene.	
KW	Cancer associated antigen; diagnosis; research; treatment; human;	
KW	breast cancer; colon cancer; gastric cancer; lung cancer;	
KW	prostate cancer; ss.	
OS	homo sapiens.	
PN	WO9904265-A2.	
PD	28-JAN-1999.	
PF	15-JUL-1998; U14679.	
PR	22-JUN-1998; US-102322.	
PR	17-JUL-1997; US-896164.	
PR	10-OCT-1997; US-061599.	
PR	10-OCT-1997; US-061765.	
PR	11-OCT-1997; US-948705.	
PR	11-OCT-1997; GB-021697.	
PA	(LUDW-) LUDWIG INST CANCER RES.	
PI	Chen Y, Gout I, Gure A, O'Hare M, Ohta Y, Old LJ,	
PI	Pfeundschuh M, Sanin U, Scanlan MJ, Stockert E,	
PI	Tureci O;	
DR	WPI: 99-132448/11.	
PT	New isolated cancer associated nucleic acids and polypeptides -	
PT	isolated using sera from cancer patients; used to develop products	
PT	for the diagnosis; monitoring or treatment of cancers	
PS	Claim 67; Page 638; 787p; English.	
CC	The invention relates to a method for diagnosing a disorder characterised	
CC	by expression of a human cancer associated antigen precursor coded for by	
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a	
CC	biological sample isolated from a subject with an agent that specifically	
CC	binds to the NAM, an expression product or a fragment of an expression.	
CC	product complexed with an HLA molecule; and (b) determining the	
CC	interaction between the agent and the NAM or the expression product as a	
CC	determination of the disorder. The products and methods can be used in-	
CC	the diagnosis, monitoring, research, or treatment of conditions	
CC	characterised by the expression of various cancer associated antigens.	
CC	The invention provides nucleic acid sequences and encoded polypeptides	
CC	which are cancer associated antigen precursors expressed in human breast	
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and	
CC	lung cancer.	
SQ	Sequence 2162 BP; 528 A; 573 C; 622 G; 439 T;	

Query Match 100.0%; Score 2162; DB 1; Length 2162;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	cctggccgcgtgcgcgtcttccaagctctgcagcggcaccgaaagaac	60
DB	1	cctggccgcgtgcgcgtcttccaagctctgcagcggcaccgaaagaac	60
QY	61	gggtctgcaacagcagctgacctggcccaagcagtgaccgaaagtggccagaa	120
DB	61	gggtctgcaacagcagcagctgacctggcccaagcagtgaccgaaagtggccagaa	120

121 ttcgagcataaagtgagatttcttgaatgtaaaatgacagagaagactatctatgat 180  
121 ttcgggcaatgaagtgagatttcttgaatgtaaaatgagagagactatctatgat 180  
181 gtgcgtggaatgtaccacagacatgagcgtgctgctgtggtgagagactgaaagctg 240  
181 gtgcgtggaatgtaccacagacatgagcgtgctgctgtggtgagagactgaaagctg 240  
241 gtacataatgaacccagcgtctgctctgtttgattgacattcgagcgtgataccactg 300  
241 gtacataatgaacccagcgtctgctctgtttgattgacattcgagcgtgataccactg 300  
301 aagcaccagctgagaaatgatcatgacgtgaccccccgctgacaggaagctgaagagtg 360  
301 aagcaccagctgagaaatgatcatgacgtgaccccccgctgacaggaagctgaagagtg 360  
361 cgtcgtgagcgtctgacacccagagcctcgagctgagtgctgctgtggtgagctgtgagttt 420  
361 cgtcgtgagcgtctgacacccagagcctcgagctgagtgctgctgtggtgagctgtgagttt 420  
421 ggtctgagcgtctgacacccagcgtgacagcgtgacagcagcgtcgagctg 480  
421 ggtctgagcgtctgacacccagcgtgacagcgtgacagcagcgtcgagctg 480  
481 caggtgaagggagcagatcgtccgagatcaatgagatatccatctcctctgtacccatgag 540  
481 caggtgaagggagcagatcgtccgagatcaatgagatatccatctcctctgtacccatgag 540  
541 gaggtcacaacctatttcgaaaccaaagaactgtgtccatcaaatgagagaaactcgag 600  
541 gaggtcacaacctatttcgaaaccaaagaactgtgtccatcaaatgagagaaactcgag 600  
601 ctgacccccgtgaaagctctctcatatgagcccccaactgtgcaatgagtgtt 660  
601 ctgacccccgtgaaagctctctcatatgagcccccaactgtgcaatgagtgtt 660  
661 gttctggaactctgaggcgtgagcagcgtgagctccccctggaactcgagaaacaag 720  
661 gttctggaactctgaggcgtgagcagcgtgagctccccctggaactcgagaaacaag 720  
721 gagagaagaggtctcatatgagcgtgagtgctccgagcgttggctgagcaattccagc 780  
721 gagagaagaggtctcatatgagcgtgagtgctccgagcgttggctgagcaattccagc 780  
781 gggcccatccagaaagcctgagcctcttcatcagacatgtaaaacgtgagcctgctgct 840  
781 gggcccatccagaaagcctgagcctcttcatcagacatgtaaaacgtgagcctgctgct 840  
841 gaggtggaattgagagatagggagacagattgtcgaagtcaatgagcgtgagccttctaac 900  
841 gaggtggaattgagagatagggagacagattgtcgaagtcaatgagcgtgagccttctaac 900  
901 ctgagtcacaagggagcgtgtaaatgtgtctgaaaaaatagccgagcgtgacatctccatt 960  
901 ctgagtcacaagggagcgtgtaaatgtgtctgaaaaaatagccgagcgtgacatctccatt 960  
961 gtagctgagcgtgagcagcgtgctgtcatcagacagcggagacggcctgacagagcggcg 1020  
961 gtagctgagcgtgagcagcgtgctgtcatcagacagcggagacggcctgacagagcggcg 1020  
1021 cagcgtgagcgtgagcagcgtgctgtcatcagacagcggcgtgagcgtgagcgtgagc 1080  
1021 cagcgtgagcgtgagcagcgtgctgtcatcagacagcggcgtgagcgtgagcgtgagc 1080  
1081 aagatccctccagggagcagcagagagtgagagcgtgagcagagagagagagagagag 1140  
1081 aagatccctccagggagcagcagagagtgagagcgtgagcagagagagagagagagag 1140  
1141 gcaagcag 1200  
1141 gcaagcag 1200  
1201 aagtttaag 1260

1201 aagtttaag 1260  
1261 accatcactgctgag 1320  
1261 accatcactgctgag 1320  
1321 tgaag 1380  
1321 tgaag 1380  
1381 ccgagctcctacagcacaag 1440  
1381 ccgagctcctacagcacaag 1440  
1441 cttccccaattggag 1500  
1441 cttccccaattggag 1500  
1501 tggtagcatgtgaaag 1560  
1501 tggtagcatgtgaaag 1560  
1561 caacctggtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 1620  
1561 caacctggtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtg 1620  
1621 cgaactgtgtgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1680  
1621 cgaactgtgtgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1680  
1681 gaagtcacaag 1740  
1681 gaagtcacaag 1740  
1741 cgtgacaacaagcctcgagcagccttgaagagagagagagagagagagagagagagagag 1800  
1741 cgtgacaacaagcctcgagcagccttgaagagagagagagagagagagagagagagagag 1800  
1801 tctctggag 1860  
1801 tctctggag 1860  
1861 ccaagaataag 1920  
1861 ccaagaataag 1920  
1921 cagccgttggag 1980  
1921 cagccgttggag 1980  
1981 aattgtgtctccag 2040  
1981 aattgtgtctccag 2040  
2041 agtctctctctccatccct 2100  
2041 agtctctctctccatccct 2100  
2101 ctccaaccttacttgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2160  
2101 ctccaaccttacttgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2160  
2161 aa 2162  
2161 aa 2162

RESULT 2  
X39644  
ID X39644 standard; DNA: 2236 BP.  
AC X39644;  
DT 02-JUL-1999 (first entry)

DE Renal cancer associated gene.  
KM Cancer associated antigen: diagnosis; research; treatment; human;  
KM breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KM prostate cancer; ss.  
OS Homo sapiens.  
PN M09904265-A2.  
PD 28-JAN-1999.  
PF 15-JUL-1998; U14679.  
PR 22-JUN-1998; US-102322.  
PR 17-JUL-1997; US-896164.  
PR 10-OCT-1997; US-061599.  
PR 10-OCT-1997; US-061765.  
PR 10-OCT-1997; US-948705.  
PR 11-OCT-1997; GB-021697.  
PA (LUDM-) LUDWIG INST CANCER RES.  
PI Chen Y, Gout I, O'Hare M, Ohta Y, Old LJ,  
PI Fireundschuh M, Sahin U, Scanlan MJ, Stockert E,  
PI Tureci O;  
DR MPI; 99-13248/11.  
PT New isolated cancer associated nucleic acids and polypeptides -  
PT isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
PS Claim 67; Page 435-436; 787pp; English.  
CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
SQ Sequence 2236 BP; 548 A; 588 C; 653 G; 447 T;

Query Match 96.1%; Score 2078; DB 1; Length 2236;  
Best Local Similarity 96.7%; Pred. No. 0;  
Matches 2162; Conservative 0; Mismatches 0; Indels 74; Gaps 1;

QY 1 cctggccggctcgagctgagctcttccagctctgagccgagccgaggaac 60  
DB 1 CCGGGCCCGGTGCGGCGGCTTTCCAGCTCCGCGAGCCGGCCACCCGAAGAAC 60  
QY 61 gggctgcaagcagcagctgagctggcccgagccatgagccgaaagtcgagaa 120  
DB 61 GGGGTGCGACAGCAGCGAGCTGAGCTGGCCAGCCATGAGCCGAAAGTGGCCGAGAA 120  
QY 121 ttccggcataagtgagatttctgatatgaatgatgcagagaagagactatcatgat 180  
DB 121 TTCGGGATTAAGTGGATTTCGTGATGAAATGATGCAAGAGACTATCTCTATGAT 180  
QY 181 gtgtctcgaaatgtacacacagacatgagcgtggtcgctgtgagagcctgaaagt 240  
DB 181 GTGCTGGAATGTATACCAACAGACCATGAGAGTGGCCGTGCTGGAGAGACTGAACCTG 240  
QY 241 gttataatgaataccagcgtgtgctctgtttgatgcaattggccgctgatacccaagt 300  
DB 241 GTATATGAATGAACCCACCTGTGCTCTGTTGATGCCATTTGGCGCTGATGCCACTG 300  
QY 301 aagcacaagtgagataatgatacagctgagcccgagctccagagaagctgaaagagt 360  
DB 301 AAGCACAGGTGGAATATGATACGTACACCCCGCGGCTCCAGAAAGCTGAAGAGAGTG 360  
QY 361 cgtctgagacgctgcaaccccgaaagcctcgagctgagtgctgtgtgtgcctgagatt 420  
DB 361 CGTCTGAGACGCTGCGACCCGAAAGGCTCGGCTGAGTGTGCTGTGAGCTGGAATTT 420  
QY 421 ggtgtggtgtcttcaatctccacactcaataaagcggtgaagcagacagctgcggctc 480

DB 421 GGCTGTGGCTCTTTCATCTCCACATCATMAAGGCGGTGAGGACAGACAGCTGGGCTC 480  
QY 481 caagtagggagcagatgcgtccgatalcatatgatalctctctctgtatccatag 540  
DB 481 CAGGTAGGGAGCAGATGCTCCGGATCAATGATATTCATCTCTCTCTACCATGAG 540  
QY 541 gaagtcataactcatctgaacccaagaacatgctgtccatcaaatgtgaacacatcgc 600  
DB 541 GAGGTATCACTCACTCATTCAGAACGAAAGAACTGTGCTCAAGAGAGACATCGGC 600  
QY 601 ctgaccccgtagaagcctctctgatatgagccctcaactgagcagatgtagatcaatt 660  
DB 601 CTGATCCCGCTGAAAGACTCTCTGATGAGCCCTCCTCATTTGGAGATGATGATTT 660  
QY 661 gtgtcggaatctggtggcgtgtcgagagcctggtgtccctcgtaaatcggaacaaag 720  
DB 661 GTGTGGAAATCTGGGGCGTGCAGAGCAGCTGGGCTGCCCTGGAAATCGGAAACAG 720  
QY 721 gaagaagaagttctcaatcagcctgtgtaggtctcccgagcctgtgtcagcatctcagc 780  
DB 721 GAGAAGAAGGTCTTCATCACCTGTAGGCTCCCGAGGCTTGCGTGCACATTTCCAGC 780  
QY 781 ggcaccatcagaagcctgtgcatcttatcagcgaatgtgaacctgtccctgtctgt 840  
DB 781 GCGCCATCCAGAAAGCTGGCATCTTTATACGCATGTGAACCTGGCTCCCTGTCTGT 840  
QY 841 gaagtgagattgtagatagagacagatgtgtgaagtcaatgtcgtcgtctctaac 900  
DB 841 GAGGTGGATTTGAGATAGAGGAGACAGATGTGTGAAGTCAATGGCGTGCATCTCTAAC 900  
QY 901 ctgtatcaagaagagctgttaaatgtgtgtaaaatagccagcctgtacatctcatc 960  
DB 901 CTGATATCAAGAGAGGCTGTAAATGTGCTGAAAATATGCGCAGCTGACCATCTCATTT 960  
QY 961 gtatgtcagctgtgcccggagctgtgtatcatagacagccggagcgtgtgagagagcgc 1020  
DB 961 GTAGCTGCACTGTGCGGAGCTGTTCATACAGACCGGAGGAGGCTGGAGAGGCGCG 1020  
QY 1021 caagctgagctgtcagcggcagagctctctcatgcaagaagcgtgtgagatgtccaac 1080  
DB 1021 CAGCGTAGCTGAGCGGCGAGAGCTTCTCATCAGAAAGCGGCTGGCGATGAGTCCAAC 1080  
QY 1081 aagatctccaagagcagcagagatgtgagcgtgcaaaagagaagaattgccagaag 1140  
DB 1081 AAGATCTCCAGAGCAGCGAGAGATGAGACGGCCAAAGGAAAGAAATTTGCCAGAAC 1140  
QY 1141 gcagcaagaagaaatgagaataccggaagagatgaaacagattttagagagagagaag 1200  
DB 1141 GCAGCAGAGAAATGAGATACCGGAAGAGATGGAACAGATTTGTAGAGAGAGAGAG 1200  
QY 1201 aagtttaagaagaatgtggaagaagactgtgggtctcaagaagaacagctactctgtctaa 1260  
DB 1201 AAGTTTAAAGAAAGCAATGGAAGAAAGACTGGGGCTCAAGAGAACAGACTCTTGCTTAA 1260  
QY 1261 accataactgtgtgagtaacccaagtaaccccttcgcaagccaagt----- 1307  
DB 1261 ACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCAAAATGTATGATCAGGGAGTG 1320  
QY 1307 ----- 1307  
DB 1321 GAACCTGAGCTCGAGCCCGCAGATGACCTTGATGAGAGCAGGAGAGAGAGACAG 1380  
QY 1307 gattccggaataatgagaagccttgaaccctactctatgttcaacccagagacatc 1366  
DB 1381 GATTTCGGAAATATGAGAGAGGCTTTGACCCCTACTCTATGTTTCAACCCAGAGAGATC 1440  
QY 1367 atgtggagaagatgtccggtctcagcatcaagaagaagagatcttaagaactgtgcctgt 1426  
DB 1441 ATGGGGAAGATGTCCGGCTCTACAGCATCAAGAGAGGAGATCTTAAAGACTGGGCTGT 1500  
QY 1427 gaagggcgtgtgagatcccatcttggaaggtgtgtctctctgtgtatagagcgggga 1486



```
Db 661 GTGTGGAAATCTGGGGGCTGCGAGGAGCGCTCCCTGGAAATCGGAAACAG 720
QY 721 gagaagaagttcttcaaccctgtgtagctcccgaggcttggcttgcacattccagc 780
Db 721 GAGAAGAAGGCTTTCACACACCCTGTGAGGCTCCCGAGGCTTGGCTTCACACATTTCAGC 780
QY 781 ggcgcccatccagaagccttgcatttatacagcccatgtgaaccttgcctcttgcct 840
Db 781 GGGCCCATCCAGAAAGCTGGCATCTTTATACGCCATGTGAACCTGGCTCCCTGTCTGCT 840
QY 841 gaagctggagattgagatagagagaccagatgttcgaagtcgaatgtgccttccctaac 900
Db 841 GAGGTGGAGATTGGAGATAGAGGAGACAGATTGTGCAAGTCAATGGCGTTCGACTTCTTAAC 900
QY 901 ctgtgatacaagaagagctgttaaatgtctgtaaaaataagccgacgcttgcacatccatt 960
Db 901 CTGATATCAAGAGAGGCTGTAAATGTCTGAAAAATAGCCGACCCGACATCTTCATT 960
QY 961 gtatgtcgaactgtgcccggagctgttcatagacagaccggagcgctgtgcagaagcgcg 1020
Db 961 GTAGCTGCACTGGCCGGAGCTGTTCTATACAGACCCGGAGCGGCTGTGCGAGAGCGCGG 1020
QY 1021 caagctgtagctgtcagcgcgagagcttctcatgtcagaagcgctgtgcagatgtccaac 1080
Db 1021 CAGGTGAGCTGCGAGCGGCGAGAGCTTCTCATGCGAAGCGGCTGGCGATGGAGTCCAAC 1080
QY 1081 aagatctccaagagcagcagagagatgtgagcgcaaaagagaaagaattgccagaag 1140
Db 1081 AAGATCTCCAGAGAGCGAGGAGATGAGACGGCGAAAGGAAATAATGCCCGCAAGAG 1140
QY 1141 gtagcagaagaaatagagaataccggagaagagatgaagaaattgtatgagagagagag 1200
Db 1141 GCAGCAGAGGAAATAGAGATACCGGAGAGAGATGGAACAGATTGTAGAGAGAGAGAG 1200
QY 1201 aagtttaagaagaatgtggaagaagacttgggtctcaaaagaaacagctactcttgcctaa 1260
Db 1201 AAGTTTAAGAAGCAATGTGGAAGAAGACTGGGGCTCAAAAGAACAGCTACTTGGCCTAAG 1260
QY 1261 accataactgtctaggttacaccacgaatacccttcgcagaccgaagt----- 1307
Db 1261 ACCATCACTGCTGAGGTACACCGAGTACCCCTTCGCAAGCAAAAGTATGATCAGAGGAGTG 1320
QY 1307 ----- 1307
Db 1321 GAACCTGAGCTCGAGCCCGGAGATGACCTGTGAGTGAGGACGACGAGAGAGAGACAG 1380
QY 1307 gattccggaataatagagaagacttgcacccactactatgttcaaccacagacagatc 1366
Db 1381 GATTTCGGAAATATAGAGAAAGGCTTGACCCCTACTCTATGTTCAACCCAGAGCANATC 1440
QY 1367 atgtggagaagatgtccggtctcctacgcatcaagaagaagagatccttagacctgtgcctg 1426
Db 1441 ATGGGGGAAGGATGTCCGGCTCTCTACGCATCAAGAGAGAGGATCTTATGAGCTGGCCCTG 1500
QY 1427 gaagggggtgtgagctcccccatttggagaaggtgtgttctgtgtgtgtgtgtgtgtgtgt 1486
Db 1501 GAAGGGGgtgtgagctcccccatttggagaaggtgtgttctgtgtgtgtgtgtgtgtgtgt 1560
QY 1487 gctgtcgaagcagatgtgtgcatgttgaagaagagagatcaatgtgaatcaacagcgaag 1546
Db 1561 GCTGCTGAGGGGATGTGTGCAATTGTCAAAAGGGGAGACAGATCATGCAATCAACGGCAAG 1620
QY 1547 attgtgaacagactaacacctgtgtgagctgtgagctgtcccttgcaagaagccttgaatcag 1606
Db 1621 ATTGTGACAACTAACACCCCTGCTGAGGCTGTGAGCTCCCTGAGAGAGGCTGGAATCAG 1680
QY 1607 gggggggagcagatgtgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1666
Db 1681 GGGGGGgagcagatgtgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1740
QY 1667 ctgacctctctgtgaagtcacaaaggggaacaaatcaacgcttgaagaaacagtgag 1726
Db 1741 CTGACCTTCTTGTGTAAGTCCAAAGGGGAACCAATTATCACCGTTAGGAACAGTGAG 1800

QY 1727 ctccggcccaactcgtgaaacacaaagctcggagccacttgaagagagccacatgaca 1786
Db 1801 CTCGGGCCCCACCTCGTGAACACAAAGCTTGAGACCGCTTGAAGAGGCGCACATGACA 1860
QY 1787 cacacagatgtgacatccttggacctgaatctatcaaccaggaatccaaaccttgg 1846
Db 1861 CACACGATGGCATCTTGTGGACCTGATCTTATCACCCAGGATCTCAATCTCCCTTGG 1920
QY 1847 gccctgaaccaaagggccagataaagaaagacgtcggccacttlttgaaggccaatgtga 1906
Db 1921 GCCCTGAACCAAGGGCCAGATTAAGAAACAGCTCGGGCCACTTTTGTGAAGCCAAATGTGA 1980
QY 1907 ggaagagggagcagcagcgcttgggagaagatctcaagaatccagaatccatctt 1966
Db 1981 GGAAGAGGAGACACCAACCTTGTGGAGAGATCTCAAGGATCCAGACTCTTCATT 2040
QY 1967 cctctggcccaatgaattgtctctccagacttgggagaccccttgccttgaacctaa 2026
Db 2041 CCTCTGGCCCAAGTGAATTTGTGTCTCTCCAGCTTTGGGGAGCTCTTCTTGAACCTTA 2100
QY 2027 taagaaccacatgtgagctctctctctcaatccctctcctctgtcccttgcctaatgtc 2086
Db 2101 TAAGACCCCACTGGAGCTCTCTCTCTCCATCCCTCTGCTGCTGCTCTGCTCTTAATTGC 2160
QY 2087 tgcagagattgtacatcccaacttactctgagctcatataataaacaagattat 2146
Db 2161 TGCAGGATTGTCACTCCAAACCTTACTGTGAGCTCATTAATAAATAACAGATTATT 2220
QY 2147 ttccagcttaaaaaa 2162
Db 2221 TTCCAGCTTAAAAAA 2236

RESULT 4
X40061
ID X40061 standard; DNA; 2289 BP.
AC X40061;
DT 02-JUL-1999 (first entry)
DE Colon cancer associated gene.
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
OS Homo sapiens.
PN W09J04265-A2.
PD 28-JAN-1999.
PE 15-JUL-1998; U14679.
PR 22-JUN-1998; US-102322.
PR 17-JUL-1997; US-896164.
PR 10-OCT-1997; US-061599.
PR 10-OCT-1997; US-061765.
PR 10-OCT-1997; US-948705.
PR 11-OCT-1997; GB-021697.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, Ohare M, Ohta Y, Old LJ,
PI Freundschun M, Sahin U, Scanlan MJ, Stockert E,
PI Tureci O;
DR WPI; 99-132448/11.
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
PS Claim 67; Page 663-664; 787pp; English.
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
```

CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
SQ Sequence 2289 BP; 567 A; 596 C; 668 G; 458 T;

Query Match 93.7%; Score 2025; DB 1; Length 2289;  
Best Local Similarity 94.5%; Pred. No. 0;

Matches 2162; Conservative 0; Mismatches 0; Indels 127; Gaps 1;

```
OY 1 cctggcccggtgctgctgctgcttcccaactcctgcaacccgggaccccggaagac 60
DB 1 cctggcccggtgctgctgctgcttcccaactcctgcaacccgggaccccggaagac 60
OY 61 gggctgctgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaac 120
DB 61 gggctgctgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaac 120
OY 121 ttcggcaataaggatgatttctgattgaaatgacagaagaagactatctatgat 180
DB 121 ttcggcaataaggatgatttctgattgaaatgacagaagaagactatctatgat 180
OY 181 gttgctgcaatgtacacacacacacacacacacacacacacacacacacacacac 240
DB 181 gttgctgcaatgtacacacacacacacacacacacacacacacacacacacacac 240
OY 241 gtcacaaatgaacacacacacacacacacacacacacacacacacacacacac 300
DB 241 gtcacaaatgaacacacacacacacacacacacacacacacacacacacacac 300
OY 301 aagcaacaggtggaatatgatcagctgaccccccgcgcgtccagaagaagtgaaggatg 360
DB 301 aagcaacaggtggaatatgatcagctgaccccccgcgcgtccagaagaagtgaaggatg 360
OY 361 cgttcggacacgtctgacacacacacacacacacacacacacacacacacacacacac 420
DB 361 cgttcggacacgtctgacacacacacacacacacacacacacacacacacacacacac 420
OY 421 ggcctggtgacacgtctgacacacacacacacacacacacacacacacacacacacacac 480
DB 421 ggcctggtgacacgtctgacacacacacacacacacacacacacacacacacacacacac 480
OY 481 caggtgaaggagacagatcgctccggaatcaatgatatccatctcctctgtaacacatgag 540
DB 481 caggtgaaggagacagatcgctccggaatcaatgatatccatctcctctgtaacacatgag 540
OY 541 gaggtcacaacactcattccgaacacgaagaactggtccatcaagaagtgaacaacatcgcc 600
DB 541 gaggtcacaacactcattccgaacacgaagaactggtccatcaagaagtgaacaacatcgcc 600
OY 601 ctgaccccggtgaaagactctcctgtaggcctcactggaactggaactggaactggaact 660
DB 601 ctgaccccggtgaaagactctcctgtaggcctcactggaactggaactggaactggaact 660
OY 661 gtttcggaatctggtggcggtgacgagacgacgacgacgacgacgacgacgacgacgacg 720
DB 661 gtttcggaatctggtggcggtgacgagacgacgacgacgacgacgacgacgacgacgacg 720
OY 721 gagaagaaggtcttcatcaagcctggtgaagctccggaagcctggtggaagccttccagc 780
DB 721 gagaagaaggtcttcatcaagcctggtgaagctccggaagcctggtggaagccttccagc 780
OY 781 gggcccaacacgaagcctggaactcttcatcaagcctggaacacggtgacgacgacgacgac 840
DB 781 gggcccaacacgaagcctggaactcttcatcaagcctggaacacggtgacgacgacgacgac 840
OY 841 gaggtggaattggaagatagggaacagattgltcgaagtcaatgagcgtggaactctcctaac 900
DB 841 gaggtggaattggaagatagggaacagattgltcgaagtcaatgagcgtggaactctcctaac 900
OY 901 ctggtacacaaagagagcgtgtaaatgtgtcgaataatagccgacgacgacgacgacgacgac 960
DB 901 ctggtacacaaagagagcgtgtaaatgtgtcgaataatagccgacgacgacgacgacgacgac 960
```

```
OY 961 gtaagtcagctggtccggagagctgttcatgacagaccgggacggtgacgaagcgcg 1020
DB 961 gtaagtcagctggtccggagagctgttcatgacagaccgggacggtgacgaagcgcg 1020
OY 1021 cagcgtgagctgcaagcggcgaagagcttctcatgcaagaacggtgctgcaatggaatccaac 1080
DB 1021 cagcgtgagctgcaagcggcgaagagcttctcatgcaagaacggtgctgcaatggaatccaac 1080
OY 1081 aagatcctcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1140
DB 1081 aagatcctcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1140
OY 1141 gcaagcaggaataatgagatcacccggaagagatggaacacattgtgaggaaggaag 1200
DB 1141 gcaagcaggaataatgagatcacccggaagagatggaacacattgtgaggaaggaag 1200
OY 1201 aagtttaagaagaacatggaagaagactggtggtcgaagaagaacagctactctgtcctaa 1260
DB 1201 aagtttaagaagaacatggaagaagactggtggtcgaagaagaacagctactctgtcctaa 1260
OY 1261 accatcactgctggaagtacacccatgaccccttgcaagccaaagt 1307
DB 1261 accatcactgctggaagtacacccatgaccccttgcaagccaaagt 1307
OY 1307 -----gattcc 1313
DB 1307 -----gattcc 1313
OY 1381 ccacagagatggttaagagatggtggtttatcaagacagcattcaagacacagattccc 1440
DB 1381 ccacagagatggttaagagatggtggtttatcaagacagcattcaagacacagattccc 1440
OY 1314 ggaataatgagaagaagcttggacccctactctatgttaaccccaagaagaatcagggga 1373
DB 1314 ggaataatgagaagaagcttggacccctactctatgttaaccccaagaagaatcagggga 1373
OY 1441 ggaataatgagaagaagcttggacccctactctatgttaaccccaagaagaatcagggga 1500
DB 1441 ggaataatgagaagaagcttggacccctactctatgttaaccccaagaagaatcagggga 1500
OY 1374 aggaatcgcgctcctcaacacgaagaagaagatctcctgaacgctggtgacgacgacgacg 1433
DB 1374 aggaatcgcgctcctcaacacgaagaagaagatctcctgaacgctggtgacgacgacgacg 1433
OY 1501 aggatcgcgctcctcaacacgaagaagaagatctcctgaacgctggtgacgacgacgacg 1560
DB 1501 aggatcgcgctcctcaacacgaagaagaagatctcctgaacgctggtgacgacgacgacg 1560
OY 1434 gttggaactccccaatggaagaattgtcgttctgctgtgtaagaagcgggagctgctg 1493
DB 1434 gttggaactccccaatggaagaattgtcgttctgctgtgtaagaagcgggagctgctg 1493
OY 1561 gttggaactccccaatggaagaattgtcgttctgctgtgtaagaagcgggagctgctg 1620
DB 1561 gttggaactccccaatggaagaattgtcgttctgctgtgtaagaagcgggagctgctg 1620
OY 1494 agcgcacatggtgacattgtgaagaaggacgaatcatatggaacacgaagatgtgga 1553
DB 1494 agcgcacatggtgacattgtgaagaaggacgaatcatatggaacacgaagatgtgga 1553
OY 1621 agcgcacatggtgacattgtgaagaaggacgaatcatatggaacacgaagatgtgga 1680
DB 1621 agcgcacatggtgacattgtgaagaaggacgaatcatatggaacacgaagatgtgga 1680
OY 1554 cagactacaacccctgctgagcgtgacgctgacgctgacgctgacgctgacgctgacgctgacg 1613
DB 1554 cagactacaacccctgctgagcgtgacgctgacgctgacgctgacgctgacgctgacgctgacg 1613
OY 1681 cagactacaacccctgctgagcgtgacgctgacgctgacgctgacgctgacgctgacgctgacg 1740
DB 1681 cagactacaacccctgctgagcgtgacgctgacgctgacgctgacgctgacgctgacgctgacg 1740
OY 1614 actgatacgaactgtgtgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1673
DB 1614 actgatacgaactgtgtgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1673
OY 1741 actgatacgaactgtgtgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1800
DB 1741 actgatacgaactgtgtgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1800
OY 1674 tcttgctgaagtccaaagggaagaacaaatccaagcgttgggaacagtgagctccggc 1733
DB 1674 tcttgctgaagtccaaagggaagaacaaatccaagcgttgggaacagtgagctccggc 1733
OY 1801 tcttgctgaagtccaaagggaagaacaaatccaagcgttgggaacagtgagctccggc 1860
DB 1801 tcttgctgaagtccaaagggaagaacaaatccaagcgttgggaacagtgagctccggc 1860
OY 1734 cccacactgctggaacacgaagcctggaacacgctggaagagggcacaatgacacacaca 1793
DB 1734 cccacactgctggaacacgaagcctggaacacgctggaagagggcacaatgacacacaca 1793
OY 1861 cccacactgctggaacacgaagcctggaacacgctggaagagggcacaatgacacacaca 1920
DB 1861 cccacactgctggaacacgaagcctggaacacgctggaagagggcacaatgacacacaca 1920
OY 1794 gatgacatccttgggaactgtatcaacccgaagaatcctcaactccttggccctga 1853
DB 1794 gatgacatccttgggaactgtatcaacccgaagaatcctcaactccttggccctga 1853
OY 1921 gatgacatccttgggaactgtatcaacccgaagaatcctcaactccttggccctga 1980
DB 1921 gatgacatccttgggaactgtatcaacccgaagaatcctcaactccttggccctga 1980
OY 1854 accaagggcagataagaagacagctcggtgacacatttttgaaggccaatgtggaaggaag 1913
DB 1854 accaagggcagataagaagacagctcggtgacacatttttgaaggccaatgtggaaggaag 1913
OY 1981 accaagggcagataagaagacagctcggtgacacatttttgaaggccaatgtggaaggaag 2040
DB 1981 accaagggcagataagaagacagctcggtgacacatttttgaaggccaatgtggaaggaag 2040
```

QY 1914 gagcagccagccgcttggagagaagatctcaagatccagactcctccttccctctg 1973  
|||||  
Db 2041 GAGCAGCCACCCGTTGGGAGAAAGATCTCAAGATCCAGACTCTCATTTCTTCTCTGG 2100  
QY 1974 cccagtgaaattgctctcccccagcttggggagctcctcttgaaacctaaagaac 2033  
|||||  
Db 2101 CCCAGTGAATTTGGTCTCTCCACAGCTTTGGGGGAGCTCTTCTGTAACCCCTAATTAAGACC 2160  
QY 2034 ccaactgagctctctctctccacccctcctcctcctcctctgcttaattgctgcag 2093  
|||||  
Db 2161 CCACTGGAGTCTCTCTCTCCATCCCTCTCTGCTCTGCTTAATTTGCTGCCAGG 2220  
QY 2094 attgtacacccaacactctctgagctcatataaataaacaagattatttcagc 2153  
|||||  
Db 2221 ATTGTCTCTCAACCTTACTCTGAGCTCATTAATAAATAACAGATTAATTATTCAGC 2280  
QY 2154 ttaaaaaa 2162  
|||||  
Db 2281 TTTAAAAAA 2289

RESULT 5  
X40062  
ID X40062 standard: DNA; 2409 BP.  
AC X40062:  
DE 02-JUL-1999 (first entry)  
KW Colon cancer associated gene.  
KW Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KW prostate cancer; ss.  
OS Homo sapiens.  
PN KC9904265-A2.  
PD 28-JAN-1999.  
PF 15-JUL-1998; 014679.  
PR 22-JUN-1998; US-102322.  
PR 17-JUL-1997; US-896164.  
PR 10-OCT-1997; US-061599.  
PR 10-OCT-1997; US-061765.  
PR 10-OCT-1997; US-948705.  
PR 11-OCT-1997; GB-021697.  
PA (LUDW-) LUDWIG INST CANCER RES.  
PI Chen Y, Gout I, Gure A, O'Hare M, O'bata Y, Old LJ,  
PI Pfeundschn M, Sahin U, Scanlan MJ, Stockert E,  
PI Tureci O.  
DR WPI; 99-132448/11.  
PT New isolated cancer associated nucleic acids and polypeptides -  
PT isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
PS Claim 67, Page 665-666; 787pp; English.  
CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
SQ Sequence 2409 BP; 595 A; 638 C; 692 G; 484 T;

Query Match 88.1%; Score 1905; DB 1; Length 2409;  
Best Local Similarity 89.7%; Pred. No. 0;  
Matches 2162; Conservative 0; Mismatches 0; Indels 247; Gaps 1;

QY 1 cctggcccgctgcgctgcgctcttcacagctcctggcagccgagcagcgaagaaac 60  
|||||  
Db 1 CCGGGCCCGGTCCGGCGGCTCTTCCAGCTCTGGACGCGGACCCGACCAAGAAC 60

QY 61 gggtctgcaacagcagcagcttgagcctcgcccaagccatgagcaggaatgagcccgagaa 120  
|||||  
Db 61 GGGTCTGCAACAGCAGCAGGTGGACCTGGCCAGCCATGGACCGAAAAAGGGCCCGGAA 120  
QY 121 ttccggcataaggttgagattcttgatgaaatgagtcagagaagagactatctatgat 180  
|||||  
Db 121 TTCCGGCATAAAGGTGGATTCTTCATTTGAAAATGATCAAGAAAGACTTCTCTATGAT 180  
QY 181 gtgtcggaaatgataccaccagacatlgagcgtgctgctgctgaggagacagcgtgaaagt 240  
|||||  
Db 181 GTGTCCGAATGTATACCAACAGACACATGAGCTGGCCCTGTCGAGGAGACCTGAAGCTG 240  
QY 241 gtaataatgaaccagcagctctgctctgtttgagtcattggccgctgatacccaact 300  
|||||  
Db 241 GTATCAATGAACCCAGCCCTCTGCTCTGTGATGCCATTTGGCGCTGATGCCACTG 300  
QY 301 aagcaccaggtggaataatgatacagctgaccccccggcgctccaggaagctgaaagagtg 360  
|||||  
Db 301 AAGCACCAGGTGGAAATGTGATCACTGACCTACCCCGCGCTCCAGAAAGCTGAAGAGAGTG 360  
QY 361 cgtctgagccgctctgcaaccaccgaagcctcgccctgagctgctgagctgagctgagctt 420  
|||||  
Db 361 CGTCTGAGACCGCTGACACCCCGAAGGCTCGGCTGAGTGCGGTGGCTGGAGAGTTT 420  
QY 421 ggcctgagctcttcaatctccacatcaaaagcgctgagcagcagcagcagcagcagctc 480  
|||||  
Db 421 GGCCTGAGGCTCTTCATCTCCACCTCATCAAAAGCGCTGAGCGAGACAGCTGCGGCTC 480  
QY 481 caagtgaggagagatgctgctcagatcaatgagatattcaatctcctctctaccatgag 540  
|||||  
Db 481 CAGGTAGGGAGCAGATGCTCCGATCAATGAGATATTCATCTCTCTTACCATGAG 540  
QY 541 gaggatcaacatctcagcaacaaagaacatgctgctcaatcaaaagtgaacacatcgcc 600  
|||||  
Db 541 GAGGTATCAACCTCATTCAGAACCAAAACATGTGTCATCAAGAGTGAACATCGCC 600  
QY 601 ctgattcccgctgaaagctctcctgataagccctcacttgagagatagtgtgagcaatt 660  
|||||  
Db 601 CTGATCCCGCTGAAAACCTCTCCGATGAGACCCCTCACTTGAGAGTATGTGATCAATG 660  
QY 661 gtgtcggaaatctggggcgctgagcagcagcctggtcctccctggaatcgaggaaataag 720  
|||||  
Db 661 GTGTGGAAATCTGGGGCGTGGCAGGCAAGCTGGGCTCCCTGGAAATCGGAAACAG 720  
QY 721 gagaagaagctctcatcaacagcctgtagctcccgagagccttgctgagcatttcagc 780  
|||||  
Db 721 GAGAAAGAGTCTTCATCAGACCTGCTAGGCTCCCGAAGCCTTGCTGACACATTTCCAGC 780  
QY 781 ggcacacccaagaagcctgcatcttatacagcatgtgaacctggtcctgctgct 840  
|||||  
Db 781 GGGCCCATTCAGAAAGCCTGGCATCTTATCAGCATGTGAACCTGCTCCCTGTCGT 840  
QY 841 gagggtgattgtaagataaggagggccagatgtgtgaagtcagatgagctgctcctcaac 900  
|||||  
Db 841 GAGGTGGATTGGAGATGAGGAGCCAGATTTGCAATGTGCAATGCGCTGCACTTCTTAAC 900  
QY 901 ctgatatcaagaagagctgtaaatgtgctgaaatatgcccagcagcagcctgacatctccat 960  
|||||  
Db 901 CTGATACAAAGGAGGCTGTAAATGTGCTGAAAAAATAGCCGACAGCTGACATCTCAT 960  
QY 961 gtatgctgagctgagccggagagctgttcaatgacagaccggagcgagctgacagagcgcg 1020  
|||||  
Db 961 GTAGCTGACAGCTGGCCCGGAGCTGTTTCATGACAGACCGGAGCGGCTGACAGAGCGCG 1020  
QY 1021 caagcgtgagctgacagcgagcgagccttctatgacagaagcgctgagatgagctccaac 1080  
|||||  
Db 1021 CAGGTGAGCTGACAGCGGAGAGCTTTCATGCAAGAGCGCTGGATGAGATCTCAAC 1080  
QY 1081 aagatcctcagagagcagcagagatgagcagcgaaggaagaaatgctcccaag 1140  
|||||  
Db 1081 AAGATCTCCAGAGAGCAGCAGAGATGAGACCGCAAGAGGAAAGAAATTTGCCCAAG 1140



QY	1141	gcagcagagagaaatctgagataccggaaagaaatggaacagattgtagagaaagag	1200
Db	1141	GCAGCAGAGGAAATGTGAGATACCCGAAAGAAATGTAACAAGTTGTAAAGAGAAAG	1200
QY	1201	aagtttaagcaatctggagaagaagacttgggtctcaaggaagaaocgtactctgtctaa	1260
Db	1201	AAGTTTAAGAACAAATGGAAAGAAAGATGGGGCTCAAAAGAACAGTACTCTTGGCTTAA	1260
QY	1261	aacatcaactgtctgaagtaaccgaataccctctcgaaagccaagt -	1307
Db	1261	ACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAACCCAAAGTATGATCAGGAGTG	1320
QY	1307	-----	1307
Db	1321	GAACCTGAGCTCGAGCCCGCAGATGACTGGATGAGGACACGAGAGACGAGGAGAGAG	1380
QY	1307	-----	1307
Db	1381	ACATTGTCACCAACCCACAGCCTCCACGAGGCCCTGGGTGTCCACCATCTCCAAACT	1440
QY	1307	-----	1307
Db	1441	GTCATGTCACACAGAGCCCAATTTCATCTACAGGCCAGTGTGAATCTGAATTCTG	1500
QY	1307	-----gattcc	1313
Db	1501	CCACGAGAGATGTTGAAGAGAGTGGTTTATCAAGACAGCATTCACACAAAGTTTTTC	1560
QY	1314	ggaataatgaagaaagctctggaacctactctatgtctcaaccocagcagatcatgaggga	1373
Db	1561	GGAATATGAGAGAGGCTTTGACCCCTACTCTATGTTCAACCCACAGACGATCATGGGGA	1620
QY	1374	aggaatgtccggtctctcagcatcaagaagaagaggaatcccttgaacctgagccctgaaagcg	1433
Db	1621	AGGATGTCCGGCTCTCTACGCATCAAGAGAGGAGGATCTTGAACCTGGCCCTGGAAGCG	1680
QY	1434	gtctgagactccccatctgggaaggtgtgtctctcgtctgtatgagcggggagctctgt	1493
Db	1681	GTGTGAGACTCCCCCATTTGGGAGAGTGGTGTCCTTCTGTATGAGCGGGAGCTGCTG	1740
QY	1494	agcgagcatgtgagatctgtgaagaagggaagaataatgcaatgaacgaagcaagatcttga	1553
Db	1741	AGCGGCATGTGGCATTTGTAAAGGGGACAGATCATGCGATCAACGGCAAGATTGTGA	1800
QY	1554	cagaactacaacctggtctgaaggtctgacgcttgccctgacagaagccttgaatcagggcggg	1613
Db	1801	CAGACTACACCTGGCTGTGAGGCTGTACGCTGCCGACGAAGGCTGTGAATCAGGGCGGG	1860
QY	1614	actgtagtcgacctgtgtgttgcgctctgcgccccaagagatatacgaatgactgacct	1673
Db	1861	ACTGGAATCGACCTGTGTGGTTGGCGTGTGCCCCCAAGGAGTATACGATGAGCTGACT	1920
QY	1674	tcttgcgtgaagtcaccaaaagggaaccaaatcaagcgtttagaaacagttagactcgcgc	1733
Db	1921	TCTTGTCTGAAGTCCAAAGGGGAAACCAATTCACCGCTTAGAANAAGTGTGCTCCGG	1980
QY	1734	cccaactcgtgaaacacaaagcctccggaacagaccttgaagaagggcacatgaacacacaa	1793
Db	1981	CCCACTCTGTGAACACAAAGCCTCGGACCGCTTGAGAGAGGCCATGACACACACA	2040
QY	1794	gatygcatactcttgggacctgaaatctatcaaccaagaaatctcaaatcccttggccttga	1853
Db	2041	GATGGAATCCTTTGGGACTCAATCTATCAACCCAGGAAATCTCAAACTCCCTTTGGCGCTGA	2100
QY	1854	accgagggcagataaagaaagcttgggcacactttttaaggccaatgttgagaaggaag	1913
Db	2101	ACCGAGGCCAGATTAAGAAAGCTCGGGCCACTTTTGAAGGCCCAATGTGAGGAAAGG	2160
QY	1914	gagcagcgaacgcttltgggagaagactcaagaatccaactctatctctctctgtg	1973
Db	2161	GAGCAGCGAACCGCTTTGGGAGAAAGATCTCAAGGATCCAACTCTCATTTCTTCTCTGG	2220
QY	1974	cccaagttaattgtctctccagcttltggggagctcctctcttgaaccttaataagac	2033

Db	2221	CCACGTGAATTTGGTCTCTCCACAGCTTTGGGGACTCTTCCCTTGAAACCTTAATATAGACC	2280
QY	2034	ccactggaagctctctctctccatccctctctctgccccttgctctaaatlgctgcagag	2093
Db	2281	CCACGTGAATTTGGTCTCTCTCCACAGCTTTGGGGACTCTTCCCTTGAAACCTTAATATAGACC	2340
QY	2094	attgtcactccaaacctctctctgagctcattaaataaacaagattatlttcagc	2153
Db	2341	ATTGTCACTCCAAACCTTACTGTGAGCTCATTAATAATAACAGATTATTTTCCAGC	2400
QY	2154	ttaaaaaa 2162	
Db	2401	TTTTTTTTTT	
Db	2401	TTTTTTTTTT	
RESULT	6		
ID	X40060	standard; DNA; 1306 BP.	
AC	X40060:		
DE	02-JUL-1999	(first entry)	
DT	Colon cancer associated gene.		
KW	Cancer associated antigen; diagnosis; research; treatment; human;		
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;		
KW	prostate cancer; ss.		
OS	Homo sapiens.		
PN	MO9804265-A2.		
PD	18-JAN-1999.		
PF	15-JUL-1998; U14679.		
PR	22-JUN-1998; US-102322.		
PR	17-JUL-1997; US-896164.		
PR	10-OCT-1997; US-061599.		
PR	10-OCT-1997; US-061765.		
PR	10-OCT-1997; US-948705.		
PR	11-OCT-1997; GB-021697.		
PA	(LUDWIG) LUDWIG INST CANCER RES.		
PI	Chen Y, Gout I, Gure A, Ohare M, Obata Y, Old LJ,		
PI	Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,		
PI	Tureci O,		
DR	WPI; 99-132448/11.		
PT	New isolated cancer associated nucleic acids and polypeptides -		
PT	isolated using sera from cancer patients, used to develop products		
PT	for the diagnosis, monitoring or treatment of cancers		
PS	Claim 67: Page 662: 787pp: English.		
CC	The invention relates to a method for diagnosing a disorder characterised		
CC	by expression of a human cancer associated antigen precursor coded for by		
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a		
CC	biological sample isolated from a subject with an agent that specifically		
CC	binds to the NAM, an expression product or a fragment of an expression		
CC	product complexed with an HLA molecule; and (b) determining the		
CC	interaction between the agent and the NAM or the expression product as a		
CC	determination of the disorder. The products and methods can be used in		
CC	the diagnosis, monitoring, research, or treatment of conditions		
CC	characterised by the expression of various cancer associated antigens.		
CC	The invention provides nucleic acid sequences and encoded polypeptides		
CC	which are cancer associated antigen precursors expressed in human breast		
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and		
CC	lung cancer		
CC	Sequence 1306 BP; 348 A; 331 C; 375 G; 252 T;		
Query Match	53.1%; Score 1148; DB 1; Length 1306;		
Best Local Similarity	94.3%; Pred. No. 3.2e-295;		
Matches 1232; Conservative	0; Mismatches 0; Indels 74; Gaps 1;		
QY	931	aaaaatagccgagcctgaccatctccattgtagctgacgctgagccggagctgttcacg	990
Db	1	AAAAATAGCCGACGCTGACCAATTCATTTGAGCTGACGTGCGCGGAGAGCTTTCCAG	60
QY	991	acagacgcggagcgcgctgacagagcgcgagcgcgtgagctgacagcgacagagctctc	1050
Db	61	ACAGACGCGGAGCGGCTTGCAAGAGCGCGGACGCTGAGCTTCACACGCGAGAGCTTTC	120

Query Match	Best Local Similarity	53.1%;	Score 1148;	DB 1;	Length 1306;
Matches 1232;	Conservative	0;	Mismatches	0;	Indels 74; Gaps 1
0y	931	aaaaatagccgagcgttacatctccatcttgatctgtagcagctgtgcgcggagctgttcacg	990		
Db	1	AAAAATAGCCCGACGAGCTGACCACTTCATTTGAGCTGCAGCTGCGCGGACGCTGTTCATG	60		
0y	991	acagaccgggagcggctgacagagcgcggcagcgttgatgagctgcaacgggaagaagcttc	1050		
Db	61	ACAGACCGGGAGCGGCTGCGCAGAGCGCGGCGAGCGTGAAGCTTCACACGCGAGGAGCTTCTC	120		







PI Masucci MG;  
DR WPI: 98-312463/27.  
PT New fusion proteins resistant to proteolytic degradation -  
PT a peptide sequence containing glycine repeats  
PS Disclosure: Fig 4b; 120pp; English.  
CC This is a nucleotide sequence of the stabilizing sequence-encoding  
CC insert. The invention provides a method for increasing the resistance  
CC of a core protein to proteolytic degradation that comprises linking or  
CC inserting onto or into the core protein a stabilizing polypeptide of  
CC formula (Gly)X(Gly)Y(Gly)Z where Glya, Glyb, Glyc are 1-6  
CC sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,  
CC Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not  
CC be identical from n repeat to n repeat. Alternatively a nucleic acid  
CC encoding the stabilizing polypeptide can be linked onto or inserted into  
CC a nucleic acid encoding a core protein. The fusion proteins of the  
CC invention are more resistant to degradation by proteases and, thus, have  
CC a longer half-life than the unfused core protein. The products can be  
CC used for treating autoimmune diseases, cancer and inflammation. In  
CC particular, the core protein may be an IkappaB regulator protein for the  
CC treatment of inflammatory bowel disease, or a nitroreductase protein  
CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer  
CC or other pathological conditions. The fusion proteins can also be used in  
CC diagnostic methods such as in vivo imaging.  
SQ Sequence 799 BP; 201 A; 106 C; 479 G; 13 T;

Query Match 2.6%; Score 57.2; DB 1; Length 799;  
Best Local Similarity 50.7%; Pred. No. 9.5e-06;  
Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 976 cggagagctgttcatacagaccggagcggtgacagagcgcgagcgctgagctgcag 1035  
DB 42 CGGAGAGCGCAGAGAGAGGCGGCGAGACAGAGAGGCGGAGAGAGCGAGCGGCG 101  
QY 1036 cggagagagcttcatacagagcggtgacagagcgcgagcgctgagctgcag 1095  
DB 102 CAGGAGGCGCAGAGAGGCGGCGAGAGAGAGAGAGGCGGCGAGAGAGAGAG 161  
QY 1096 cagcagagagatggagcgcaagagagaagaatgtccagaaggcagcagagaaat 1155  
DB 162 GGGCAGAGGCGCAGAGAGGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGG 221  
QY 1156 gagaggtaccggaagagatgtgacagatgtgtgagagagagaagttaagaagcaa 1215  
DB 222 GAGGGGCGAGAGAGAGAGAGGCGGCGAGAGAGGCGGCGAGAGAGAGAGAGG 281  
QY 1216 tgggaagaagacttgggtctcaagaagacag 1245  
DB 282 CAGGAGCAGAGAGAGAGGCGGCGAGAGAGGCGGCGAG 311

RESULT 10  
V21683  
ID V21683; standard; DNA; 9600 BP.  
AC V21683; (first entry)  
DT 17-AUG-1998  
DE Vector plasmid pCMVkmITR-EPI.  
KW Polynucleotide delivery; vaccine; polycationic agent; ss.  
OS Chimeric - Epstein-Barr virus.  
OS Chimeric - Adeno-associated virus.  
OS Chimeric - Cytomegalovirus.  
OS Chimeric - Bos taurus.  
FH Location/Qualifiers  
FT Key 14..2594  
FT CDS  
FT /tag= a  
FT /product= "EBV nuclear antigen A"  
FT /tag= b  
FT /note= "EBV origin of replication"  
FT repeat\_unit 4928..5104  
FT /tag= c

FT /rpt\_type= INVERTED  
FT /note= "AAV inverted terminal repeat"  
FT repeat\_unit 7189..7355  
FT /tag= d  
FT /rpt\_type= INVERTED  
FT /note= "AAV inverted terminal repeat"  
FT promoter 5112..6734  
FT /tag= e  
FT /note= "CMV immediate-early enhancer/promoter"  
FT terminator 6818..7050  
FT /tag= f  
FT /note= "bovine growth hormone polyA sequence"  
PN WO9806437-A2.  
PD 13-FEB-1998.  
PF 13-AUG-1997; 014465.  
PR 13-AUG-1996; US-023867.  
PA (CHIR ) CHIRON CORP.  
PI Cohen F, Dubois-Stringfellow N, Dworki V, Innis MA, Murphy JE, Tetsuo U, Zukermann R;  
DR WPI: 98-139296/14.  
PT Polycationic agents based on alpha-amino acids, able to complex  
PT with nucleic acid - to facilitate its entry into cell, condense it  
PT and protect it against serum degradation, particularly for use in  
PT gene therapy  
PS Disclosure: Page 77-80; 100pp; English.  
CC This polynucleotide comprises the DNA sequence of vector plasmid  
CC pCMVkmITR-EPI, which contains an Epstein-Barr virus (EBV) origin  
CC of replication from plasmid pCEP4, a coding region for EBV nuclear  
CC antigen A from pCEP4, a pair of inverted terminal repeats from  
CC adeno-associated virus, a cytomegalovirus enhancer/promoter, a  
CC bovine growth hormone polyA sequence, and a kanamycin resistance  
CC selectable marker. Polynucleotides encoding polypeptides, such as  
CC erythropoietin or leptin, and ribozymes and antisense  
CC polynucleotides can be inserted into the vector. The vector is  
CC preferred for use in novel compositions and methods for improved  
CC polynucleotide delivery into cells. In these methods, polycationic  
CC agents are used to increase the frequency of uptake of a  
CC nucleic acid (see also V21684-86) into a cell. The polycationic  
CC agent can condense with the nucleic acid and inhibit serum and/or  
CC nuclease degradation of the nucleic acid. The nucleic acid can be  
CC a vector, may express a therapeutic protein or a vaccinating viral  
CC or cancer antigen, or is itself therapeutic (antisense or  
CC ribozyme). The methods and compositions can be used in the gene  
CC therapy of many diseases.  
SQ Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T;

Query Match 2.6%; Score 57.2; DB 1; Length 9600;  
Best Local Similarity 50.7%; Pred. No. 3.3e-05;  
Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 976 cggagagctgttcatacagaccggagcggtgacagagcgcgagcgctgagctgcag 1035  
DB 713 CGGAGAGCGCAGAGAGAGGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGG 772  
QY 1036 cggcagagagcttcatacagagcggtgacagagcgcgagcgctgagctgcag 1095  
DB 773 CAGGAGGCGCAGAGAGGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGG 832  
QY 1096 cagcagagagatggagcgcaagagagaagaatgtccagaaggcagcagagaaat 1155  
DB 833 GGGCAGAGGCGCAGAGAGGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGG 892  
QY 1156 gagagataccggaagagatgtgacagatgtgtgagagagagaagttaagaagcaa 1215  
DB 893 GAGGGGCGAGAGAGAGAGGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGG 952  
QY 1216 tgggaagaagacttgggtctcaagaagacag 1245  
DB 953 CAGGAGCAGAGAGAGAGGCGGCGAGAGAGGCGGCGAG 982

RESULT 11

ID	051731
AC	051731 standard; DNA; 10596 BP.
DT	31-MAY-1994 (first entry)
DE	Plasmid pCISBON for subcloning hHGF variants.
KW	Hepatocyte Growth Factor; HGF; variant; mutain; in vitro mutagenesis
KW	proteolysis resistant; liver; malignancy; CMV-driven;
KW	Cytomegalovirus; episomal expression plasmid; ss.
OS	Synthetic.
FH	Key
FT	enhancer
FT	/tag= a
FT	/note= "CMV enhancer/promoter"
FT	promoter
FT	758..775
FT	/tag= b
FT	/label= SP6_promoter
FT	845..849
FT	/tag= c
FT	/note= "SP6 RNA start"
FT	misc.feature
FT	902..966
FT	/tag= d
FT	/function= cloning_linker
FT	967..1107
FT	polya-signal
FT	/tag= e
FT	/note= "SV40 poly A"
FT	misc.feature
FT	1108..1531
FT	/tag= f
FT	/function= SV40_origin
FT	1580..4189
FT	/tag= g
FT	/label= EBNA-1
FT	4190..6374
FT	/tag= h
FT	/function= orip
FT	4295..4887
FT	repeat_region
FT	/tag= i
FT	/note= "family of repeats"
FT	misc.structure
FT	5866..5978
FT	/tag= j
FT	/note= "dyad region"
FT	6375..6457
FT	/tag= k
FT	/label= HSV-TK_terminator_3'-end
FT	cds
FT	6975..7975
FT	/tag= l
FT	/phenotype= neomycin_resistance
FT	/note= "n5 neomycin phosphotransferase gene"
FT	promoter
FT	7975..8112
FT	/tag= m
FT	/label= TK_Promoter
FT	8114..8594
FT	/tag= n
FT	/function= M13_ori
FT	misc_rna
FT	8595..10414
FT	/tag= o
FT	/label= delta_2a
PN	WO9323541-A.
PD	25-NOV-1993.
PE	17-MAY-1993: U04648.
PR	18-MAY-1992; US-884811.
PR	18-MAY-1992; US-885971.
PA	(GETH ) GENENTECH INC.
PI	Godowski PJ, Lokker NA, Mark MR;
DR	WPI: 93-386573/48.
PT	Hepatocyte growth factor variants - are resistant to proteolytic cleavage into its two-chain form, used to treat malignancies associated with HGF receptor
PS	Example 1; Fig 6; 87pp: English.
CC	Plasmid pCISBON (a pRK5 derivative) is an episomal CMV driven expression plasmid. hHGF variants with enhanced receptor binding activity were produced by site-directed mutagenesis. Stable populations of preferred HGF variants were obtained by transfecting human embryonic kidney 293 cells and then these were subcloned in

[illegible]

RESULT 12  
 T40348  
 T40348 standard; DNA; 10596 BP.  
 AC T40348;  
 DT 09-DEC-1996 (first entry)  
 DE Plasmid pcisBON for expression of hepatocyte growth factor.  
 KW Human; hepatocyte growth factor; HGF; hHGF; serum; proteolytic cleavage  
 KW pro-hormone; beta subunit; alpha subunit; kringle domain; prothrombin;  
 KW plasminogen; catalytic domain; serine protease; HGF variant;  
 KW HGF receptor; malignancy; chronic HGF receptor activation; ss.  
 OS Synthetic.  
 PN US547856-A.  
 PD 20-AUG-1996.  
 PF 18-MAY-1992; 884811.  
 PR 18-MAY-1992; US-885971.  
 PR 18-MAY-1992; US-884811.  
 PR 13-JUL-1993; US-087783.  
 PA (GERTH ) GENENTECH INC.  
 PI Godowski, PJ, Locker NA, Mark MR;  
 PI WPI; 96-392634/39.  
 PT New hepatocyte growth factor variants - are resistant to in vivo  
 PT proteolytic cleavage into a 2-chain form, useful as HGF antagonists  
 PS Example 1; Fig 6; 39pp; English.  
 CC This sequence represents the episomal CMV driven expression plasmid  
 CC pcisBON which was used in the expression of variant human hepatocyte  
 CC growth factor (HGF). HGF is isolated from human serum and is a  
 CC disulphide linked heterodimer derived by proteolytic cleavage of the  
 CC pro-hormone between residues 494 and 495. This generates a molecule  
 CC composed of an alpha subunit of 440 amino acids (mol. wt. 69 kd) and  
 CC a beta subunit of 234 amino acids (mol. wt. 34 kd). The alpha and beta  
 CC subunits are encoded by a single open reading frame. The alpha subunit  
 CC contains four kringle domains based on their homology to kringle-like  
 CC domains in other proteins, e.g. prothrombin, plasminogen. The beta  
 CC subunit shows high homology to the catalytic domain of serine proteases.  
 CC However two of the three residues which form the catalytic triad of  
 CC serine proteases are not conserved in HGF. Therefore, the precise  
 CC function of the beta chain remains unknown. The invention includes HGF  
 CC variants which retain HGF receptor binding activity without having the  
 CC biological activity of wild-type HGF. They can be used for the treatment  
 CC of pathological conditions associated with the activation of a HGF  
 CC receptor such as malignancies associated with chronic HGF receptor  
 CC activation. The pcisBON plasmid comprising the HGF coding sequence  
 CC may be used for manipulation of the HGF coding sequence and expression  
 CC of the variant HGF's of the invention.

Seq	Sequence	10596 BP:	2627 A:	2571 C:	3023 G:	2375 T:
Query Match		2.6%;	Score 57.2;	DB 1;	Length 10596;	
Best Local Similarity		50.7%;	Pred. No. 3.5e-05;			
Matches 137;	Conservative	0;	Mismatches 133;	Indels	0;	Gaps
0;						
QY	976	cggagagctggttcatatgacagaccggyagcgctgtgcagagagcgcgagcgcgtgtagctgcag	1035			
DB	2247	cgggagggggcagagacagagagggggcagagagcagagagagaggggagagagagagagagag	2306			
OY	1036	cggcagagagagcttctcatatcagaagagcggtcgtggagatggaggtccaaagaatccttcagag	1095			
DB	2307	cagcagggggcagagagcggcagagagggggcagagagcagagagagagagagagagagagag	2366			
OY	1096	cagcagagagatggagcggcgaagaggaagaaatgtgtcccaagagcagcagagagaaat	1155			
DB	2367	ggcgacagagagggcagagagggcagagagcagagagagagagagagagagagagagagagag	2426			
OY	1156	gagagataccggaagagagatggaacagatgttaagaggaggaagaaagttaagaagcaa	1215			
DB	2427	gagggggcagagacagagagagagggcagagagggcagagagggcagagagagagagaggg	2486			
OY	1216	tggaagaagaactgggggtcctaagaagacag	1245			
DB	2487	cagcagcagagagagagagggcagagagggcag	2516			
RESULT	13					
ID	X15650					
AC	X15650	standard; DNA; 10596 BP.				
DT	10-MAY-1999	(first entry)				
DE	Nucleotide sequence of plasmid pCIS.EBON.					
KW	Plasmid pCIS.EBON; hepatocyte growth factor; HGF; variant;					
OS	HGF receptor activation; ss.					
PN	Synthetic.					
PD	09-MAR-1999.					
PF	09-FEB-1994; 194087					
PR	09-FEB-1994; US-194087.					
PA	(GETH ) GENENTECH INC.					
PI	Godowski PJ, Lokker NA, Mark MR;					
DR	WPI; 99-203949/17.					
PT	New hepatocyte growth factor variants - useful for studying					
PT	structure-function relationships in the wild type molecule and for					
PT	treating conditions associated with chronic hepatocyte growth factor					
PT	receptor activation					
PS	Example 1, Fig 6A-F; 40pp; English.					
CS	The present sequence represents the nucleotide sequence of pCIS.EBON,					
CC	which is used in the course of the invention. The specification					
CC	describes a hepatocyte growth factor (HGF) variant (HGfV) comprising an					
CC	amino acid (aa) alteration at or adjacent to position 692 of the					
CC	wild-type human HGF (huHGF) aa sequence. HGfV may be used in studies to					
CC	identify the structure-activity relationships of HGF in order to identify					
CC	the functionally important domains in the aa sequence. It may also be					
CC	used to identify aa residues which are responsible for the interaction of					
CC	HGF with its receptor, and those aa that are responsible for the					
CC	biological activity of it. Variants of HGfV which have enhanced receptor					
CC	binding affinity (compared to wild-type huHGF) and are more biologically					
CC	active than wild type huHGF, may be used as huHGF agonists. Conversely,					
CC	variants of HGfV which have enhanced receptor binding affinity (compared					
CC	to wild-type huHGF) but which are biologically inactive may be used as					
CC	huHGF antagonists, and may be used to block the binding of wild-type					
CC	huHGF to its receptor. This permits the treatment of pathogenic					
CC	conditions associated with the activation of an HGF receptor, such as					
CC	malignancies associated with chronic HGF receptor activation.					
CC	HGfV-immunoglobulin chimeras may be produced (by standard methods) and					
CC	used in protein A purification, immunohistochemistry, and					
CC	immunoprecipitation techniques (in place of anti-HGF antibodies) or in					
CC	screening studies to identify inhibitors of HGF-HGF interactions.					
Q0	Sequence	10596 BP;	2627 A;	2571 C;	3023 G;	2375 T;

Query Match	Best Local Similarity	Score	DB 1	Length
Match 137; Conservative	50.7%	Pred. No. 3,5e-05;	0; Mismatches 133;	Indels 0; Gaps 0
Qy 976	cggagagctgttcacatgcacaccgagacccggtcgcgttcggacagagccggtcagcgtgtgagctcag	1035		
Db 2247	CGGGAGGGGCGACGAGACGAGAGGGGCGACGAGACGAGAGAGAGGGCGACGAGACGAGAGGGG	2306		
Qy 1036	cgcgcagagcttcctcattcgtcagaagcgcgttcgtcgatctgagatccacaagaatccctccagga	1095		
Db 2307	CAGGAGGGGCGACGAGAGGGGCGAGAGGGGCGACGAGACGAGAGAGGGGCGACGAGAGGAG	2366		
Qy 1096	cagcagagagatgtgagcgccgcaagagagaaaagaattgtcccaagaagcagcagaggaatat	1155		
Db 2367	GGCGACGAGAGGGGCGAGAGGGGCGACGAGACGAGAGAGGGGCGACGACGACGAGAGGGGCG	2426		
Qy 1156	gagagatcccggaagagagatgagaaacagattgtagagagagagaaagttaagaagca	1215		
Db 2427	GAGGGCGACGAGACGAGAGAGGGGCGACGAGAGGGGCGACGAGAGGGGCGACGAGAGGGGCG	2486		
Qy 1216	tgggaagaagacttggggctcaagaagcaag	1245		
Db 2487	CAGGAGCGACGAGAGGGGCGACGAGGGGCGAG	2516		
RESULT 14				
ID V55830	standard; DNA; 795 BP.			
AC V55830;				
DT 18-NOV-1998	(first entry)			
DE FGA insert stabilising polypeptide encoding DNA.				
KM Fusion protein; stabilising polypeptide; proteolytic degradation;				
KM resistance; half-life; autoimmune disease; inflammation; nitro drug;				
KM Ikappab regulator protein; inflammatory bowel disease; in vivo imaging;				
KM nitroreductase protein; enzyme therapy; prodrg therapy; protease;				
KM cancer; pathological condition; ss.				
OS Epstein-barr virus.				
PH Key	Location/Qualifiers			
FT CDS	1..788			
FT	/*tag= a			
FT	/product= "stabilising polypeptide"			
PN MO9822577-A1.				
PD 28-MAY-1998.				
PF 17-NOV-1997; IB1508.				
PR 25-JUN-1997; US-048945.				
PR 15-NOV-1996; US-030986.				
PA (MASU/) MASUCCI M G.				
PI Masucci MG;				
DR WPI: 98-312463/27.				
DR P-PSDS: W79128.				
PT New fusion proteins resistant to proteolytic degradation -				
PT comprising a core protein with a stabilising polypeptide comprising				
PT a peptide sequence containing glycine repeats				
PS disclosure; Fig 3; 120pp; English.				
CC This DNA encodes a stabilising polypeptide and is the FGA insert of the				
CC invention. The invention provides a method for increasing the resistance				
CC of a core protein to proteolytic degradation that comprises linking or				
CC inserting onto or into the core protein a stabilising polypeptide of				
CC formula [(Gly)nX(Glyb)(Glyc)Z]n where Glya, Glyb, Glyc are 1-6				
CC sequential Glya residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,				
CC Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not				
CC be identical from n repeat to n repeat. Alternatively a nucleic acid				
CC encoding the stabilising polypeptide can be linked onto or inserted into				
CC a nucleic acid encoding a core protein. The fusion proteins of the				
CC invention are more resistant to degradation by proteases and, thus, have				
CC a longer half-life than the unfused core protein. The products can be				
CC used for treating autoimmune diseases, cancer and inflammation. In				
CC particular, the core protein may be an Ikappab regulator protein for the				
CC treatment of inflammatory bowel disease, or a nitroreductase protein				
CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer				
CC or other pathological conditions. The fusion proteins can also be used in				
CC diagnostic methods such as in vivo imaging.				



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 19, 2000, 22:33:48 ; Search time 3256.04 Seconds  
(without alignments)  
2691.327 Million cell updates/sec

Title: US-09-502-945-5  
Perfect score: 2162  
Sequence: 1 ccctggcccgctgcgctgcg.....tatttcagcttaaaaaa 2162

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.	Score	Match Length	DB	ID	Description
C 1	578.2	26.7	616	61	AL045824
C 2	477.2	22.1	512	79	AM300030
C 3	457.8	21.2	473	74	AI265960
C 4	440.4	20.4	485	39	AA860926
C 5	437.6	20.2	467	51	AI761354
C 6	435.4	20.1	464	51	AI739238
C 7	423	19.6	473	63	AI984995
C 8	405.6	18.8	636	26	W22160
C 9	403	18.6	411	49	AI628985
C 10	403	18.6	411	49	AI629007
C 11	394.2	18.2	407	63	AI989344
C 12	372.8	17.2	384	60	AI796741
C 13	371.2	17.2	385	44	AI272081
C 14	371.2	17.2	465	80	AM390711
C 15	369.2	17.1	386	46	AI453367
C 16	354.2	16.4	360	63	AI981419
C 17	351.4	16.3	367	79	AM304132
C 18	341.4	15.8	381	80	AM351654
C 19	335.6	15.5	342	51	AI738967
C 20	334	15.4	342	51	AI788958
C 21	304.6	14.1	311	35	AA551799
C 22	292.2	13.5	317	32	AA344913
C 23	290.4	13.4	335	28	AA111174
C 24	287.8	13.3	316	74	AM192301
C 25	270.4	12.5	272	44	AI270726
C 26	267.4	12.4	289	32	AA348243
C 27	266.4	12.3	278	63	AM013904
C 28	248.8	11.5	279	61	AI868991
C 29	240.8	11.1	244	60	AI783811
C 30	197.4	9.1	216	61	AI868907
C 31	180.8	8.4	206	31	AA337309
C 32	160	7.4	204	28	AA121198
C 33	143	6.6	353	49	AI642234
C 34	131.2	6.1	221	64	AM062449
C 35	120.6	5.6	473	36	AA616807
C 36	105	4.9	463	105	AO320634
C 37	68.2	3.2	823	84	B03396
C 38	67.2	3.1	443	102	AQ100902
C 39	58.2	2.7	597	82	CNS005E
C 40	57.2	2.6	587	36	C79507
C 41	56.6	2.6	571	92	AO919110
C 42	55.8	2.6	576	36	C79500
C 43	55.6	2.6	445	34	AA464787
C 44	55.2	2.6	498	50	AI676456
C 45	54.6	2.5	657	80	AW369742

## ALIGNMENTS

<p>Clone from S. Wiemann, Molecular Genome Analysts, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by BMF within the CDNA sequencing consortium of the German Genome Project.</p> <p>r1 sequence also available.</p> <p>This clone is available at the RZPD in Berlin.</p> <p>Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.</p>				
FEATURES	Location/Qualifiers			
source	1..616			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="DKFZp434H236"			
	/clone_1lb="434 (synonym: htes3)"			
	/tissue_type="testis"			
	/dev_stage="adult"			
	/lab_host="DH10B"			
	/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"			
BASE COUNT	134 a	142 c	185 g	155 t
ORIGIN				
Query Match	26.7%	Score 578.2;	DB 61;	Length 616;
Best Local Similarity	98.4%;	Pred. No. 2,1e-136;		
Matches	605; Conservative	0; Mismatches	8; Indels	2; Gaps
OY	1529	atgcgaatcaacggaagaattgacagaactacacctgctgagtgctagcgtgcctg	1588	
Db	613	ATGGCAATCAACGGCAAGATTGTGACAGACTACACCTCGTGAGGCTGACGCTGCCGTG	554	
OY	1588	cagaagcgctgtaatccaaggcggggaactgatacgacattgtgttgccgtctgcgcccca	1648	
Db	553	CAGAAAGCCTGGAAATCAGGCGGGGACTGTGATCGACCTTGTGGTGGCCCTGTGCCCCCA	494	
OY	1649	aaggagatgaagatgagctgacctctctgtcctgtaagtccaagaagggaaccaaatlca	1708	
Db	493	AAGGAGTATGACGATGAGCTGACCTTCTT-CGTAAGTCCAAAAGGGGAACCAAAATTCC	435	
OY	1709	gcgttagaagaacagtgaagctccggcccacactcgtgaaacaagaagccttgagcagcctt	1768	
Db	434	-CGTTAGGAAACATGAGAGCTCCGGGCCCACTCCGTGAACACAAAGCCTTGATCAGCCTT	376	
OY	1769	gagagagggccacatgacaacaccagatggacatcccttggagctgaatcatatcaccag	1828	
Db	375	GAGAGAGGCCACACTACACACACACCGATGCGACTTCTTGGGAGCTGTGATTAATCACCCAG	316	
OY	1829	aatctcaactcccttggcccttgaaaccaaggcgccagataaggaacaagctcggcaccctt	1888	

RESULT	1
LOCUS	AL0045824/c
DEFINITION	AL0045824 616 bp mRNA EST 29-SEP-1999
ACCESSION	DKEZD434H236.g1_434 (synonym: hnes3) Homo sapiens CDNA clone
VERSION	DKEZD434H236.3, mRNA sequence.
KEYWORDS	AL0045824
SOURCE	AL0045824.1 GI:5433935
ORGANISM	EST. human.
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 616)
AUTHORS	Koehler,K., Beyer,A., Mewes,H.W., Gassenhuber,J. and Wiemann, TTLLE
JOURNAL	EST (Koehler, et al.) Unpublished (1999)
COMMENT	On May 18, 1998 this sequence version replaced gi:3137188.

Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 3' sequence of the clone insert

RESULT 2  
AW300030/C

```

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de, sequenced by BKFZ within the cDNA sequencing consortium of the German Genome Project.
ri sequence also available.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14055 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1. 616
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP434H236"
/clone_id="434 (synonym: htes3)"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; site_1: NotI; site_2: SalI"
134 a 142 c 185 g 155 t

ch 26.7%; Score 578.2; DB 61; Length 616;
1 Similarity 98.4%; Pred. No. 2, 1e-136;
60; Conservative 0; Mismatches 8; Indels 2; Gaps

ctggcaatcaacgycgaagattgttacagatatacccttgctgagctgacgtgcctg 1588
ttggcattcaacggcagattgtgtacagactacacccttgctgagctgacgtgcctg 1587
ttggcattcaacggcagattgtgtacagactacacccttgctgagctgacgtgcctg 1586
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1644
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1643
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1642
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1641
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1640
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1639
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1638
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1637
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1636
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1635
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1634
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1633
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1632
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1631
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1630
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1629
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1628
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1627
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1626
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1625
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1624
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1623
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1622
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1621
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1620
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1619
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1618
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1617
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1616
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1615
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1614
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1613
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1612
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1611
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1610
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1609
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1608
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1607
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1606
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1605
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1604
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1603
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1602
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1601
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1600
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1599
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1598
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1597
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1596
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1595
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1594
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1593
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1592
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1591
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1590
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1589
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1588
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1587
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1586
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1585
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1584
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1583
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1582
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1581
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1580
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1579
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1578
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1577
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1576
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1575
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1574
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1573
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1572
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1571
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1570
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1569
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1568
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1567
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1566
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1565
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1564
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1563
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1562
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1561
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1560
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1559
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1558
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1557
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1556
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1555
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1554
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1553
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1552
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1551
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1550
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1549
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1548
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1547
aagaagcctcgggaatcagcgcgcggtgctgacatgcacctgtgttcgctcgtccccc 1546
aagaagcctcgggaatcagcgcgcg
```





Matches 459; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Oy 654 tcaatttctgtcgaatcctggggcgctgagagcagcctggctccctggaaatcggga 713
    |||
Db 461 ttagtttgggtggaatctggggcgctgagagcagcctggctccctggaaatcggga 402
Oy 714 aaacaagagaaagagcttctatcagcctgtgtaagctcccaaggcctgtgctcagcat 773
    |||
Db 401 AAACAAGGAAGAAGGCTTCATCAGCCTGTAGAGCTCCAGAGCCTTGCTGCAGCAT 342
Oy 774 ttccagcgccccatccagaagcctggcatcttcatcagcatgtgaaacctgtgctcct 833
    |||
Db 341 TTCAGCGCCGCCCATCCAGAACCTGGCATCTTTATCACCCATGTGAACCTGGCTCCCT 282
Oy 834 gctcctgagtgaggattgagatagggagaccagattgtcgaatcgaatgagctgact 893
    |||
Db 261 GCTCTCTGAGGTGGATTGGAGATAGGGGACCAATTGTCCGAATGATGGCTGCACATT 222
Oy 894 ctctaactgtgatacaagaagagctgttaattgtctgaaataagccgacgctgacat 953
    |||
Db 221 CTCTAACCTGGATCACAGAGAGGCTGTAAATGTGCTGAAGATAGCCGACCTGACCAT 162
Oy 954 ctccattgtagctcagctgtgcccggagagctgttcatgacagacgggagcgctggcaga 1013
    |||
Db 161 CTCCATTTGATGCTGACGCTGGCCGGAGCTGTTCATGACAGACCGGAGCGCTGGCAGA 102
Oy 1014 ggcgggagcgctgagctgcaacggcagagagcttctatgacagagcgagctggcagatga 1073
    |||
Db 101 GCGCGGCGACGCTGAGCTGACGCGGACGAGAGCTTCTCATGACGAAGCGGCTGGCATGGA 42
Oy 1074 gtccaacaagatcctcaggaagcagagagatgagcgcc 1114
    |||
Db 41 GTCGAACAAGATCTCCAGAGACAGACAGATGAGCGGC 1
RESULT 4
AA860926/c 485 bp mRNA EST 04-JAN-1999
DEFINITION ak22d06.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1406699
ACCESSION AA860926.1 GI:2953066
VERSION AA860926.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 485)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TUMOR Gene Index
Unpublished (1997)
JOURNAL On Jan 19, 1998 this sequence version replaced gi:2287585.
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILMIL at:
www-bio.llnl.gov/bdrp/image/image.html
FEATURES
source 1..485
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1406699"
/location="Soares_testis_NHT"
```

```
/sex="male"
/lab_host="DH10B"
/notice="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5',
TGTTCACCAATCTGAAGTGGAGCGGCGGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cots, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

Query Match 20.4%; Score 440.4; DB 39; Length 485;

Best Local Similarity 98.1%; Pred. No. 2e-101;

Matches 477; Conservative 0; Mismatches 6; Indels 3; Gaps 3;

```
Oy 1666 gctgaccttctgtcgaagtcacaaagggaacaaattcagcgcttagaagaacagcga 1725
    |||
Db 485 GCTGACCTTCTT-CTGAAGTCCAAAAGGGGAACCAAAATTCAC-CGTTAGGAACAGTGA 428
Oy 1726 gctccgccccaccctcgtgacacaaagcctcgacacagccttgagagcgacatgac 1785
    |||
Db 427 GCTCCGGCCCCACCTCGTGACACAAAGCCTTGACACACCTTGAGAGAGCGCACATAC 368
Oy 1786 acacacagatgacatccttggagctgaatctatcacccaggaatctcaactccctt 1845
    |||
Db 367 ACACACCAAGATGACATCCTTGAGACCTGAATCTACCCAGAGATCTCAAACTCCCTTT 308
Oy 1846 ggccttgaaacagcgagcagataagaaagcctcgagccactttttgaagcgcaatggg 1905
    |||
Db 307 AG-CCTGAACCAAGGCGCCAGATPAAAGAACAGCTCGGGCCACTCTTGTGAAGGCCAATGTGG 249
Oy 1906 aggaagaggaagcagcagccgcttggagagagatcgaagatccagagctcactcctt 1965
    |||
Db 248 AGGAAGAGGAGACGACGACCGCTTTGGAGAGATCTCAAGAGATCCAGACTCTCATTTCTT 189
Oy 1966 tcccttgcccagtgaaattgctctcccaagccttggggagctccttctcttgaacccta 2025
    |||
Db 188 TCCCTCGGCCAGTGAATTTGGTCTCTCCAGCTCTGGGGGAGACCTCTTGAACCCCTTA 129
Oy 2026 ataagaccccatgagctctctctccatccctcctcctcctcctcctcctcaattg 2085
    |||
Db 128 ATAAAGACCCCACTGAGAGCTCTCTCCACACCTCTCTCGCCCTGCTCTTAATTG 69
Oy 2086 ctgcccagattgcactccaaccttactctgagctcatataaataaacaagattat 2145
    |||
Db 68 CTGCCAGATTGTCTACTCCAAACCTTACTCTGAGCTCATTAATAAATAAACAGATTAT 9
Oy 2146 ttcca 2151
    |||
Db 8 TTTCCA 3
RESULT 5
A1761354/c 467 bp mRNA EST 24-JUN-1999
LOCUS A1761354
DEFINITION w160e08.x1 NCI_CGAP_Col16 Homo sapiens cDNA clone IMAGE:2394662 3',
ACCESSION A1761354.1 GI:5177021
VERSION A1761354.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 467)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
```

JOURNAL  
Unpublished (1997)  
On May 18, 1998 this sequence version replaced gi:138298.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/dbp/image/image.html

FEATURES  
source  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1. 467  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="239462"  
/clone\_lib="NCI-CGAP\_Col6"  
/tissue\_type="colon tumor, RER+"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT73D-Pac (pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Col6 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneids 1057416-1061255, and 1144584-1145351).  
Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 108 a 96 c 143 g 120 t  
ORIGIN

Query Match 20.2%; Score 437.6; DB 51; Length 467;  
Best Local Similarity 98.7%; Pred. No. 9,9e-101;  
Matches 462; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 1691 aggggaaccaatttcacgcgttaggaagaagtcgcgcgcacccctcgtgaacaca 1750  
|||||  
DB 467 AGGGGAACCAATTTCAC -CGTTAGCAAAAGTAGCTCCGCCCACTCTGTAACACA 409  
QY 1751 aagcctcgagccagccttgagagagccacatgacacacacagatggcattcgagac 1810  
|||||  
DB 408 AAGCCTCGAGCCAGCCTTGAGAGGCCACATACACACACAGATGCAATCGTTGGGAC 349  
QY 1811 ctgaatctatcacccaggaatctcaaatcccttggccttggaaccagggccagataag 1870  
|||||  
DB 348 CTGAATCTATCACCCAGGATCTCAAACTCCCTTGGCCCTGAAACGAGGCAATGAG 289  
QY 1871 aacagctcgggacact-ttttgaaggccaattgagagaaaggagcagccgctt 1929  
|||||  
DB 288 AACAGCTGGGGCACTTTTGAAGCCAAATGGAGGAAAGGAGACCAAGCCGTTT 229  
QY 1930 gggagaagatctcaagatccagacttcaatcttctccttggccaggaatttgctc 1989  
|||||  
DB 228 GGGAGAAGATCTCAAGATCCAGACCTCATTCCTTCCCTGGCCCAAGTAATTTGTC 169  
QY 1990 tctcccaagcttggggagcctctccttggaaccctaagaagcccaatgagctctc 2049  
|||||  
DB 168 TCTCCAGGCTCTGGGGACCTCTCCCTTGAACCTTAATTAAGACCCCACTGGAGTCTCT 109  
QY 2050 ctctcaatcctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 2109  
|||||  
DB 108 CTCTCAATCCCTCTCCTGCTGCTGCTCTAATTCCTGCAAGATGTCATCTCAAC 49  
QY 2110 ttactctgagctcaataaataaacaagattatttccagcttaa 2157  
|||||  
DB 48 TTACTGTGAGCTCATTAATAAATAACAGATTATTTTCCAGCTAAA 1

RESULT 6  
A1739238/c 464 bp mRNA EST 20-DEC-1999  
LOCUS w14907.x1 NCI-CGAP\_Col6 Homo sapiens cDNA clone IMAGE:2390268 3',  
DEFINITION mRNA sequence.  
ACCESSION A1739238  
VERSION A1739238.1 GI:5101219  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 464)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Unpublished (1997)  
Tumor Gene Index  
JOURNAL  
On Jun 5, 1998 this sequence version replaced gi:1389064.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/dbp/image/image.html

FEATURES  
source  
Insert Length: 582 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 421.  
Location/Qualifiers  
1. 464  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="2390268"  
/clone\_lib="NCI-CGAP\_Col6"  
/tissue\_type="colon tumor, RER+"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT73D-Pac (pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Col6 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneids 1057416-1061255, and 1144584-1145351).  
Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 104 a 90 c 140 g 130 t  
ORIGIN

Query Match 20.1%; Score 435.4; DB 51; Length 464;  
Best Local Similarity 98.7%; Pred. No. 3,6e-100;  
Matches 439; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1718 aacatgtagcttcggcccccctcgtgaacacaaagctcgacacagccttgagagagc 1777  
|||||  
DB 464 AACAGTAGCTCCGGCCCCACCTCGTAACCAAAAGCCTGGACCAAGCTTGAAGAGGC 405  
QY 1778 cacatgacacacacacagatgcatccttggagactgatatcacaccagaaatcctcaa 1837  
|||||  
DB 404 CACACTGACACACACACAGATGGATCCTTGGAGCTGATATATCACCAAGATCTCAAA 345  
QY 1838 ctcccttggccctgaacacagggccagataagaacagctcgggccaattttgaagc 1897  
|||||  
DB 344 CTCCTTTGGCCCTGGAACCGAGGCGAGATGAAGAACAGCTCGGGCCACTCTTGAAGGC 285  
QY 1898 caatgtaggaaggaaggagcagccgcttggggaagaaatcgaaggtcgaagctc 1957  
|||||

	Db	284	CATGTGAGGAAAGGGAGACGACCACCGCTTTTGGAGAAGATCTCAAGGATCCAGACTCT	225
Qy	1958	cattcccttcctcctggcccagatgaatttggctctctccaggctttggtgggaactccctt	2017	
Db	224	CAATTCCTTCCTCCTGGCCCCAGTGAAATTGGTCTCTCCAGACTCTGGGGGAACCTCTTCCT	165	
Qy	2018	gaaccctaataagaccacacgtgagctctctctctccacacctctccctgcctgccttcgc	2077	
Db	164	GAAACCTTAATMAAGACCCCAGCTGGAGTCTCTCTCTCTCTCCATCCCTCTCTCTGGCCCTCTGC	105	
Qy	2078	tctaattgtcgcagagattgtcacctccaaccttactctgagctcatataataaataaac	2137	
Db	104	TCTAATTTGGTCGCCAGGATTGTCACTCCACAACCTTACTCTGAGCTCATTTATAATAAAC	45	
Qy	2138	agatttatlttccagcttaaaaaa	2162	
Db	44	AGATTATTTTCCAGCTAAAAAAA	20	
RESULT	7	A1984995	473 bp	mRNA EST 31-AUG-1999
LOCUS	A1984995/c	wr89g06.x1 NCL_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2494906	3'	
DEFINITION	A1984995	mRNA sequence.		
ACCESSION	A1984995	GI:5812272		
VERSION	EST.			
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
REFERENCE	Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
AUTHORS	1 (bases 1 to 473)			
TITLE	NCL-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.			
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
COMMENT	Unpublished (1997)			
	On May 18, 1998 this sequence version replaced gi:3138171.			
	Contact: Robert Strausberg, Ph.D.			
	Tel: (301) 496-1550			
	Email: Robert.Strausberg@nih.gov			
	Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.			
	Emmert-Buck, M.D., Ph.D.			
	cDNA Library Preparation: M. Bento Soares, Ph.D.			
	cDNA Library Arrayed by: Greg Lennon, Ph.D.			
	DNA Sequencing by: Washington University Genome Sequencing Center			
	Clone distribution: NCL-CGAP clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LLNL at:			
	www.bio.llnl.gov/bbrp/image/image.html			
FEATURES	Seq primer: -40UP from Gibco			
source	High quality sequence stop: 463.			
	Location/Qualifiers			
	1..473			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone_image="IMAGE:2494906"			
	/clone_lib="NCL_CGAP_Kid11"			
	/lab_host="DH10B"			
	/note="Organ: kidney; Vector: pTR3D-Pac (Pharmacia) with			
	a modified polylinker; Site.1: Not I; Site.2: Eco RI;			
	Plasmid DNA from the normalized library NCL_CGAP_Kid3 was			
	prepared, and ss circles were made in vitro. Following HAV			
	purification, this DNA was used as tracer in a subtractive			
	hybridization reaction. The driver was PCR-amplified cDNAs			
	from a pool of 5,000 clones made from the same library			
	(cloneds 1322376-1323911, 1456007-1456775, and			
	1500352-1502855). Subtraction by Bento Soares and M.			
	Fatima Bonaldo."			
BASE COUNT	110 a	96 c	145 g	121 t
ORIGIN				1 others
Query Match	19.6%;	Score 4.23;	DB 63;	Length 473;

Best Local Similarity		96.0%	Pred. No. 5, 1e-97:	Matches 455, Conservative 0, Mismatches 16, Indels 3, Gaps 2,	
Qy	1677	tcgtgaagtcacaaaggggaaacccaattcaacgcttgaagaaacagtgagtcgcggccc	1736		
Db	471	ttctgaagtcacaaaggggaaacccaattcaacgcttgaagaaacagtgagtcgcggccc	413		
Qy	1737	acctgtgtaacacaaagccttcgaccagccttgagaggaagccacatgtaacacacagat	1796		
Db	412	acctgtgtaacacaaagccttcgaccagccttgagaggaagccacatgtaacacacagat	355		
Qy	1797	ggcatccttgggaacacgtaattctatacccccaggaaatcctaactcccttggccctgaac	1856		
Db	354	ggcatccttgggaacacgtaattctatacccccaggaaatcctaactcccttggccctgaac	295		
Qy	1857	agggacaga laaggaacacagctcggcgacccttlttgaagggcaatgttggaggaagagag	1916		
Db	294	agggacaga laaggaacacagctcggcgacccttlttgaagggcaatgttggaggaagagag	235		
Qy	1917	caagcagccgttggaggaagaagatctcaagatccagaatctatctccttcgtgcgc	1976		
Db	234	caagcagccgttggaggaagaagatctcaagatccagaatctatctccttcgtgcgc	175		
Qy	1977	atgtgaattgagctctcccaagcttggggagatcctctcttgaacccaataagaacc	2036		
Db	174	atgtgaattgagctctcccaagcttggggagatcctctcttgaacccaataagaacc	115		
Qy	2037	ctggagctctctctctccatccctctcctcctcgcgcctcgtcctaattgctcagagat	2096		
Db	114	ctggagctctctctctccatccctctcctcctcgcgcctcgtcctaattgctcagagat	55		
Qy	2097	gtcaactcaaaccttaactctgagctcaatataaataaacaagattattcc	2150		
Db	54	gtcaactcaaaccttaactctgagctcaatataaataaacaagattattattcc	1		
RESULT	8				
LOCUS	W22160	636 bp	mRNA	EST	06-MAY-1996
DEFINITION		63A6 Human retina cDNA Tsp5091-cleaved	sublibrary	Human sapiens	cdNA
ACCESSION	W22160				
VERSION	W22160.1	GI:1299007			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE					
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
TITLE	Eutheria; Primates; Catarrhini; Hominiide; Homo.				
JOURNAL	1 (bases 1 to 636)				
COMMENT	Macke, J., Smallwood, P. and Nathans, J.				
	Adult Human Retina cDNA				
	Unpublished (1996)				
	On Apr 14, 1993 this sequence version replaced gi:785820.				
	Contact: Dr. Jeremy Nathans				
	Dr. Jeremy Nathans, Dept. of Molecular Biology and Genetics				
	Johns Hopkins School of Medicine				
	725 North Wolfe Street, Baltimore, MD 21205				
	Tel: 410 955 4678				
	Fax: 410 614 0827				
	Email: jeremy_nathans@jhu.edu				
	Clones from this library are NOT available.				
	PCR Primers				
	FORWARD: CTTTGGACAGATTCAGCGTGGTTAGT				
	BACKWARD: GAGTGCGCTATGAGTATTTCTTCCAGGGTAA				
	Seq primer: GGCTAAAGCAAAAGATT.				
FEATURES					
SOURCE					
	Location/Qualifiers				
	1..636				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone_lib="Human retina cDNA Tsp5091-cleaved sublibrary"				
	/sex="mixed (males and females)"				
	/tissue_type="retina"				
	/dev_stage="adult"				



REFERENCE  
1 (bases 1 to 411)  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Mar 16, 1998 this sequence version replaced gi:2961764.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/MLN at:  
[www-bio.lnl.gov/dbtrp/image/image.html](http://www-bio.lnl.gov/dbtrp/image/image.html)

FEATURES  
source

Seq primer: -40UP from G1bco.  
Location/Qualifiers  
1..411  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2285306"  
/clone\_lib="NCI CGAP\_Ki611"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Ki613 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneids 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 95 a 83 c 131 g 102 t  
ORIGIN

Query Match 18.6%; Score 403; DB 49; Length 411;  
Best Local Similarity 98.8%; Pred. No. 5.8e-92;  
Matches 406; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1734 cccactcgtgaacacaaagcctcgaccagcttgagagagagcacaatgacacaca 1793  
|||||  
411 CCCACCTCGTGAACACAAAGCCTCGACCAAGCTTGAGAGAGGCCACACTACACACACCA 352  
1794 gatgcatccttgagacatgaatcatcaccaggaatctcaactccttgagccctga 1853  
|||||  
351 GATGCATCTCTGGGACCTGATCTATCATCCAGGAATCTCAAACTCTTGCGCCCTGA 292  
1854 accagggccagaataagacagctcgagccactcttttgaaggccaatgtgaggaag 1913  
|||||  
291 ACCAGGGCCAGATAAGAAAGCTCGGGCCACTCTTGAAAGGCCAATGTGAGGAAAGG 232  
1914 gagcgagcagccggttggagagaagatccaagatccagacactatctctctctgg 1973  
|||||  
231 GAGCGACCAAGCGTTGGAGAAAGATCTCAAGATCCAGACTCTATTCCTCTCTGCG 172  
1974 cccaatgaattgtctctccagctcttgaggagactcctctctgaacctataagacc 2033  
|||||  
171 CCCACTGAATTTGGTCTCTCCAGCTCTGGGGAGACTCCTCTCTTAACCTTAATAAGAC 112  
2034 ccaactgagctctctctccatccctctcctctgcacctgtcctaattgctgcagg 2093  
|||||  
111 CCACAGGAGTCTCTCTCTCCATCCCTCTCTGCGCCCTGTGCTAATGCTCTCCAGG 52  
2094 attgcatctcaaacctactctgtgctcatataataaacagattta 2144  
|||||  
51 ATTGTACTCCAAACCTTACTCTGAGCTCATTAATAATAAAGATTTTA 1

RESULT 11  
LOCUS A1989344/C  
DEFINITION w227e03.x1 NCI-CGAP\_Brn53 Homo sapiens cDNA clone IMAGE:2559292 3',  
mRNA sequence.  
ACCESSION A1989344  
VERSION A1989344.1 GI:5836225  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE  
1 (bases 1 to 407)  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BTGAAP), Tumor Gene Index  
Unpublished (1998)  
On Feb 18, 1999 this sequence version replaced gi:4296517.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life  
Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The  
I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/MLN at:  
[www-bio.lnl.gov/dbtrp/image/image.html](http://www-bio.lnl.gov/dbtrp/image/image.html)

FEATURES  
source

Seq primer: -40UP from G1bco.  
Location/Qualifiers  
1..407  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2559292"  
/clone\_lib="NCI CGAP\_Brn53"  
/tissue\_type="three pooled meningiomas"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies."

BASE COUNT 97 a 81 c 127 g 102 t  
ORIGIN

Query Match 18.2%; Score 394.2; DB 63; Length 407;  
Best Local Similarity 98.0%; Pred. No. 9.9e-90;  
Matches 399; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1744 gaacacaagcctcgaccagccttgagagagagccacatgacacacacagatgcatcc 1803  
|||||  
407 GAACACAAGCCTCGACATCAGCTTGAGAGAGGCCACACTACACACAGATGCAATCC 348  
1804 ttggagacctgaattcatatccacaggaatctcaactcctttggccctgaacaggcca 1863  
|||||  
347 TTGGGACCTGAATCTATACCCAGGAATCTCAAACTCCCTTGGCCTGAAACAGGGCCA 288  
1864 gataaggaagcgtcgagccactcttttgaaggccaatgtgaggaaggaagcagccag 1923  
|||||  
287 GATAAGGAAGAGCTCGGGCACTCTTGGAAGGCCAAGCTGAGGAGGAAGGAGCCAGC 228  
1924 ccglttggagaagatctcaagatccagactctatctcctctctctgagccagtgat 1983  
|||||  
227 CCATTGGGAGAGATCTCAAGAGATCCAGACTCTCATTCCTCTCTGCGCCCAATGAAT 168  
1984 ttgctctccagcttggaggagactcctcttbaaccttaataagcccccagagagt 2043  
|||||  
167 TTGGTCTCTCCAGCTCTGGGGAGACTCTCTCTTGAACCTTAATAAGACCCCACTGGAGT 108  
2044 ctctctctccatccctctcctctgtccctctgtcctaattgtctgcagattgtcactc 2103

Db	107	CTCTCTCTCATCCCTCTCTCTCTGCGCCCTTGCGCTTAATGTCCTCCAGCATTTGTCATC	48
QY	2104	caaaccttactctgcagctcaatataaataaacagattatttcc	2150
Db	47	CAAACTTACTCTGAGCTCATTAATAATAAACAAGATTATTATTC	1
RESULT	12		
LOCUS	AI796741/c		
DEFINITION	AI796741	384 bp mRNA	EST 16-DEC-1999
ACCESSION	AI796741	NCI_CGAP_Kid11 Homo sapiens	cdna clone IMAGE:2298051 3',
VERSION	AI796741.1	GI:5362204	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 384)		
JOURNAL	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .		
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) On May 18, 1998 this sequence version replaced gi:3138695. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution Information can be found through the I.M.A.G.E. Consortium/ILM, at: <a href="http://www-bio.illn.gov/bhrp/image/image.html">www-bio.illn.gov/bhrp/image/image.html</a>		
FEATURES	Insert Length: 496 Std Error: 0.00 Seq primer: -400P from Glpco.		
SOURCE	Location/Qualifiers 1..384 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2298051" /clone_lib="NCI_CGAP_Kid11" /lab_host="DH10B" /note="Organ: kidney; Vector: pT7n3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneds 1323376-1333911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo.		
BASE COUNT	95 a 75 c 120 g 94 t		
ORIGIN			
Query Match	17.2%; Score 372.8; DB 60; Length 384;		
Best Local Similarity	98.2%; Pred. No.2.6e-84;		
Matches 377; Conservative	0; Mismatches 7; Indels 0; Gaps 0;		
QY	1772	agaagccacatgacacacacagatgcatccttggagctgacatcatcaccagaat	1831
Db	384	AGAGGCCACACATACACACACAGATGCACTCTGGAGCTGAATCTATCACCCAGAAAT	325
QY	1832	ctcaactcccttggccctgacacagggccagataaagaagaagctggcgcaactttt	1891
Db	334	CTCAAACTCCCTTTTGCCCTTGAAACAGAGGCCAGATTAAGAAAGAGCTCGGGCACACTTCT	265

QY	1892	gaagggccaatgtgagggaggaaggaagcagccgcttgggagaagaatcctaagatcca	1951
DB	264	GAAGGCCAAGCTGGAGGAGAAAGGAGACAGCCACCAATTGGGAGAGATCTCAAGATCCA	205
QY	1952	gactcattcccttccctctgagccaccagatgttggctctctccagacttggggagactcc	2011
DB	204	GACCTCTCAATTCCTTTCCTCTGGCCAGAGAAATTTGGTCTCTCCAGCTCTGGGGAGCTCC	145
QY	2012	ttccttgaacccctaataagaaccacatggagctctctctctcactccctctctcgc	2071
DB	144	TTCCCTTGAACCCCTAATAAGACCCCACTGGAGCTCTCTCTCCATCCCTCTCCCTCC	85
QY	2072	ccctctctaattgtctgcagagattgttactcccaaccttactctcagactcattatcaa	2131
DB	84	CTCTCTCTAATTTGCTGTGCCAGGATGTCTACTCCAAACCTTACTCTGAGCTATTATATAA	25
QY	2132	ataaacagattatttccagctt	2155
DB	24	ATAAACAGATTATTATTTCCAGCTT	1
RESULT	13		
LOCUS	AI272081/c		
DEFINITION	AI272081	385 bp	mRNA
ACCESSION	AI272081	q178f10.x1	NCI_CGAP_Ki43 Homo sapiens cDNA clone IMAGE:1865611 3', mRNA sequence.
VERSION	AI272081	GI:3891248	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
COMMENT	Unpublished (1997)		
	On Jan 14, 1998 this sequence version replaced gi:1797673.		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.		
	Emmert-Buck, M.D., Ph.D.		
	cDNA Library Preparation: M. Bento Soares, Ph.D.		
	cDNA Library Arrayed by: Greg Lennon, Ph.D.		
	DNA sequencing by: Washington University Genome Sequencing Center		
	Clone distribution: NCI-CGAP clone distribution information can be		
	found through the I.M.A.G.E. Consortium/MLML at:		
	www-bio.llnl.gov/dbrrp/image/image.html		
FEATURES			
source	Insert Length: 568	Std Error: 0.00	
	Seq primer: -40UP from Gibco		
	High quality sequence stop: 379.		
	Location/Qualifiers		
	1..385		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606		
	/clone_image="IMAGE:1865611"		
	/clone_id="NCI_CGAP_Ki43"		
	/lab_host="DH10B"		
	/note="Organ: kidney; Vector: pT7p3D-Pac (Pharmacia) with		
	a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st		
	strand cDNA was primed with a Not I - oligo(dT) primer,		
	double-stranded cDNA was ligated to Eco RI adaptors		
	(Pharmacia), digested with Not I and cloned into the Not		
	I and Eco RI sites of the modified pT7p3 vector. mRNA		
	source: 2 pooled kidneys. Library went through one round		
	of normalization. Library constructed by Bento Soares and		
	M. Fatima Bonaldo.		
BASE COUNT	94 a	72 c	118 g
ORIGIN			101 t







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 00:31:43 ; Search time 243.1 Seconds  
(without alignments)  
2969.166 Million cell updates/sec

Title: US-09-502-945-2

Perfect score: 2885  
Sequence: 1 ggaattcctctgtcgaagt.....aaacaaaaagttaaattt 2885

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2885	100.0	2885	1 X40055	Colon cancer assoc
2	84.6	2.9	32207	1 V73805	KSHV LTR DNA (nucl
3	84.6	2.9	137507	1 V19941	KSHV Long unique c
4	79	2.7	799	1 V55831	Nucleotide sequenc
5	79	2.7	9600	1 V21683	Vector plasmid PCM
6	79	2.7	10596	1 051731	Plasmid pcisEBON f
7	79	2.7	10596	1 T40348	Plasmid pcisEBON f
8	79	2.7	10596	1 T40348	Plasmid pcisEBON f
9	71.2	2.5	795	1 V55830	Nucleotide sequenc
10	68.4	2.4	2004	1 V87399	FLCA insert stabl
11	66.4	2.3	424	1 V87399	Nephila clavipes s
12	65	2.3	1032	1 084832	EST clone CD146. N
13	63.8	2.2	1032	1 023646	Spinocebellar at
14	62	2.1	2338	1 021883	Acetylcholinester
15	62	2.1	2338	1 V23249	N. clavipes draglin
16	60.4	2.1	2744	1 098470	Nephila clavipes s
17	60.2	2.1	543	1 023092	Mispl.-containing p
18	59.6	2.1	117213	1 V62176	Antigen tc-7a gene
19	59	2.0	1505	1 055750	HSV-2 strain SB5 C
20	58.8	2.0	1771	1 084834	Genomic clone G11F
21	58.8	2.0	195	1 084831	Spinocebellar at
22	58.8	2.0	2000	1 N71065	Gene encoding Plas
23	58.4	2.0	5661	1 V21518	Rattus norvegicus
24	57	2.0	168	1 084833	Spinocebellar at
25	56.8	2.0	1995	1 014184	N. clavipes draglin
26	56.8	2.0	1995	1 V23250	Nephila clavipes s
27	55.6	1.9	2784	1 T65102	Luciferase gene un
28	55.4	1.9	2214	1 V22682	New DNA sequence 1
29	55.4	1.9	3331	1 V22683	New DNA sequence 1
30	55	1.9	916	1 T771743	Human V3 loop HIV
31	55	1.9	1052	1 T27712	Human pp32 cDNA. N
32	54.8	1.9	2462	1 V35475	Human hsk3 coding
33	54.6	1.9	203	1 V17226	SCA2 gene CAG repe
34	54.6	1.9	203	1 V30271	Glutamine rich reg

35	54.6	1.9	3076	1 V43674	Receptor type tyro
36	53.6	1.9	12001	1 076213	HSV L/ST region. H
37	53.2	1.8	165	1 V30274	Glutamine rich reg
38	53.2	1.8	1005	1 X15318	DNA encoding a the
39	52.8	1.8	2818	1 T10554	Sheep mammary glan
40	52.6	1.8	15377	1 025975	MH mutant porcine
41	51.8	1.8	506	1 084804	Spinocebellar at
42	51.8	1.8	3366	1 084803	Spinocebellar at
43	51.6	1.8	456	1 T69166	Trypanosoma cruzi
44	51.4	1.8	15672	1 010613	Rianodin receptor
45	51.2	1.8	2301	1 V20445	Human c-trlk oncoge

ALIGNMENTS

RESULT	1
AC	X40055 standard; DNA; 2885 BP.
AC	X40055
DE	02-JUL-1999 (first entry)
DE	Colon cancer associated gene.
KW	Cancer associated antigen; diagnosis; research; treatment; human;
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW	prostate cancer; ss.
OS	Homo sapiens.
PN	W09904265-A2.
PD	28-JAN-1999.
PF	15-JUL-1998; U14679.
PF	22-JUN-1998; US-102332.
PR	17-JUL-1997; US-869164.
PR	10-OCT-1997; US-061589.
PR	10-OCT-1997; US-061765.
PR	11-OCT-1997; US-948705.
PA	(LUDM-) LUDMIG INST CANCER RES.
PI	Chen Y, Gout I, Gure A, O'hare M, O'bata Y, Old LJ,
PI	Pfreundschuh M, Sahn U, Scanlan MJ, Stockert E,
PI	Tureci O;
DR	WPI: 99-132448/11.
PT	New isolated cancer associated nucleic acids and polypeptides -
PT	isolated using sera from cancer patients, used to develop products
PT	for the diagnosis, monitoring or treatment of cancers
PS	Claim 67, Page 651-652; 787bp; English.
CC	The invention relates to a method for diagnosing a disorder characterised
CC	by expression of a human cancer associated antigen precursor coded for by
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC	biological sample isolated from a subject with an agent that specifically
CC	binds to the NAM, an expression product or a fragment of an expression
CC	product complexed with an HLA molecule; and (b) determining the
CC	interaction between the agent and the NAM or the expression product as a
CC	determination of the disorder. The products and methods can be used in
CC	the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens.
CC	The invention provides nucleic acid sequences and encoded polypeptides
CC	which are cancer associated antigen precursors expressed in human breast
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC	lung cancer.
CC	Sequence 2885 BP; 626 A; 901 C; 838 G; 520 T;
SO	
Query Match	100.0%; Score 2885; DB 1; Length 2885;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2885; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 ggaattcctctgtcgaagtcagaagcagccacacagcgcgtcaccattccctcc 60
DB	1 ggaattcctctgtcgaagtcagaagcagccacacagcgcgtcaccattccctcc 60
OY	61 acagacaccccaaatcgtcgggagccacacatgctcttgaccagagttccctccca 120
DB	61 acagacaccccaaatcgtcgggagccacacatgctcttgaccagagttccctccca 120

QY	121	gagcgcgccccccttggaagcgctccctccctcaaaactgccttgctggccctacgaag	180
Db	121	gagcgagccccccttgggagccgctccctccctcaaaactgccttgctggccctacgaag	180
QY	181	tcgagagacactccccctccgcaaaacagccctctgaacccaacttgaagtcgcttaag	240
Db	181	tcgagagacactccccctccgcaaaacagccctctgaacccaacttgaagtcgcttaag	240
QY	241	gctaaacagaaggttgctctgaagcggaagcagtcacctccctcgcgcgcgaagga ttggac	300
Db	241	gctaaacagaaggttgctctgaagcggaagcagtcacctccctcgcgcgcgaagga ttggac	300
QY	301	tgattattagcactcttttaagaaagacacttttgagatcAACAGGTGCGGGCGGCTC	360
Db	301	tgattattagcactcttttaagaaagacacttttgagatcAACAGGTGCGGGCGGCTC	360
QY	361	gtccggttgttaacggcgaccggcgctccggcccgagctctcccaacagctcccaagpac	420
Db	361	gtccggttgttaacggcgaccggcgctccggcccgagctctcccaacagctcccaagpac	420
QY	421	catcgctgaagatgctcttaacttgctctcaagtcaccaacatccccaacttgatctgctcc	480
Db	421	catcgctgaagatgctcttaacttgctctcaagtcaccaacatccccaacttgatctgctcc	480
QY	481	gcaccgagccctccctctgagcagctccccaacagttcaagcctctaacgctctccctc	540
Db	481	gcaccgagccctccctctgagcagctccccaacagttcaagcctctaacgctctccctc	540
QY	541	ctctgccaacactctcccttaaggtctgaagcgcaagtcacgtgcacaaactaaacccac	600
Db	541	ctctgccaacactctcccttaaggtctgaagcgcaagtcacgtgcacaaactaaacccac	600
QY	601	tgccctccccaagctgtctcgacagcagcaggagccgagagcagccctccagtcctgcg	660
Db	601	tgccctccccaagctgtctcgacagcagcaggagccgagagcagccctccagtcctgcg	660
QY	661	gcagaggtgacagctgacgcggaagttcatagcacatcccttatctcgtgctgctgct	720
Db	661	gcagaggtgacagctgacgcggaagttcatagcacatcccttatctcgtgctgctgctg	720
QY	721	ggcggtgacactgtagggcgagcggggcccccaaggcaatgctctccctgcgtgagagatg	780
Db	721	ggcggtgacactgtagggcgagcggggcccccaaggcaatgctctccctgcgtgagagatg	780
QY	781	gctctgctgtaggaagcccgcgagcagagcaacccctcatctgtctgacactccacgycga	840
Db	781	gctctgctgtaggaagcccgcgagcagagcaacccctcatctgtctgacactccacgycga	840
QY	841	gtccccaactagtgacgycgtgagcgtgtgycaccaagcatlgcgagcgttaagcagctccc	900
Db	841	gtccccaactagtgacgycgtgagcgtgtgycaccaagcatlgcgagcgttaagcagctccc	900
QY	901	ggcgatctgggcccccttgccgcaactaagttccctaacgcgtgcgcgaagttccccaagccct	960
Db	901	ggcgatctgggcccccttgccgcaactaagttccctaacgcgtgcgcgaagttccccaagccct	960
QY	961	gcagcagctgtctcatgcaaacacacgacacacgacgacttcttggaagaaagcagacacaca	1020
Db	961	gcagcagctgtctcatgcaaacacacgacacacgacgacttcttggaagaaagcagacacaca	1020
QY	1021	gctaacagcttggaagatctccctcaaccaagacagggaagctgcccagcagcccaaccacca	1080
Db	1021	gctaacagcttggaagatctccctcaaccaagacagggaagctgcccagcagcccaaccacca	1080
QY	1081	ccctgagagagacagagagagagcttgacggagcagcagagagtccttgctgggggaagagac	1140
Db	1081	ccctgagagagacagagagagagcttgacggagcagcagagagtccttgctgggggaagagac	1140
QY	1141	ccctaacctgccccggagagagctcccaagaagatgtgaagcacacagaggaagaccttgagga	1200
Db	1141	ccctaacctgccccggagagagctcccaagaagatgtgaagcacacagaggaagaccttgagga	1200
QY	1201	ggagagacagagaaagagatgtgggagagagagagagatgtgcatccaggttaaggaacagga	1260

Db	1201	GGAGACGAGGAMAGCATGGGAGGAGGAGGAGATTCATCCAGGTTAAGACGAGGA	1260
Qy	1261	ggcgagagtggtgtctcgaagagggccgcgacttggagagccctgtgtctgatacaaaa	1320
Db	1261	GGGGAGAGTGTGTCTGTGAGGAGGGGCCCGACTTTGGAGGAGCGCTGTGTCTGATACAAAAA	1320
Qy	1321	actgttctcagaatgcacaacccgtcgaacctttgcaagtgatccaagggccctcagct	1380
Db	1321	ACTGTTCTCAATATCCCAACCGCTGGCAACCTTTGCAAGGTATCCAAAGGGCCCTCACCT	1380
Qy	1381	ggcaactgtgcccccaacgaccttggccgtatccaacatctccctctgtcccttgggg	1440
Db	1381	GGCCACTGTGCCCAACCCCTTGCGGTACCCATCTCTCCCTGTGCCCTGGGGGG	1440
Qy	1441	catgaagacccccccagaccacccgttcaagacctcttaccacaagtggtgtctcaga	1500
Db	1441	CATGAAGAACCCCCCAACCCACCCCTCAAGACACTCTTACACCAACGTGTGTCTACGA	1500
Qy	1501	caacttaatgttaagacaacgaatgcattgtgtggaaacaaacgctgcaccccttagacatgc	1560
Db	1501	CACGTCTATGTATTAAGACCACTGATGTGTGCGGAGACACACGCTGCACCTTGACATGC	1560
Qy	1561	tgccgcgataccaagacatctgtgtccggcgctgcagagagaacaagcctgtctaaagtgca	1620
Db	1561	TGCCCGATTCAGAGCAATCTGTGTCCGGCGTGCAGAGACAGGCTGTTCACAAGTGCGA	1620
Qy	1621	ggcgatccggaggtgtgcacaagccacgctagatgataccaagaatgcaccttgatbcca	1680
Db	1621	GCGGATCCGAGGTGTGCAAACCCACCGTATGATGATCCAGACATGCACTCTGAATACCA	1680
Qy	1681	caacctgtctatbtgagccagctcccttcaaccgycagaagctcagaacgaagaagtgtct	1740
Db	1681	CACCTGTCTATBTGAGCAAGTCCCTCCCAACCGCAGAGAGTACACACAAGAAGTTGCT	1740
Qy	1741	cggttccatcagccagaagaatgatgatgtctgtctgtcctgttggggcatccggggtggacag	1800
Db	1741	CGGTCCATCAGCCAGAAAGTGTATGCTGTGCTGTGTGGGGCATCCGGGGTGGACAG	1800
Qy	1801	tgaacccgtgtgtgaatagaatgcacctcccaagtgtgtgtgcacatggcagctggcgtct	1860
Db	1801	TGACACCGTGTGGAAATGAAATGCATCTCTCAATGCTGTGGCCATGGCAATGGGCTGCT	1860
Qy	1861	gctgtgagcttggtccttcaaggctgtgcacgaagagctcagaatgtgatttgcataatccg	1920
Db	1861	GCTGTGAGCTGTGCTTCAAGGTGGCTGTGAGAGAGCTCAAGAAATGATTTGGCATATCCG	1920
Qy	1921	ggccccaggaacacacgcgcgaagaatccacaagcattcgtgattcttcttcaactcgt	1980
Db	1921	GCCCCAGGAACACGCCCGAGGAATCCACAGCAATGGGATTTCTCTTTCACACTCTGT	1980
Qy	1981	agccaatcccgcaaaactctctacacagaagaattgaaacgtgtggcagaagcttccatcgttga	2040
Db	1981	AGCCATCAACCGCAAAACTCTTACAGCAGAAAGTTGAACGTTGGGCAAGTCTCATATGTGA	2040
Qy	2041	ctgtggacattcaatcgaatgacatgtgacacccaagcagcggttctcaatgtaccctctgtgt	2100
Db	2041	CTGTGACATTTCAACATGTGCATATGGCAACCCAGCGGTTCTTCAATGACCCCTCTGTGCT	2100
Qy	2101	ctaacatctctgtatcgtatgtacaaacgggaactcttccagagctcttgggctctctga	2160
Db	2101	CTAACATCTCTGTGATGTGCTATATGCAACGSGAATCTTTCAGGCTCTGGGGCTCTCGA	2160
Qy	2161	agaagttgtgtgaagaccagcgvtbvggttcaaatgttgaacgtgtgcacatgaacagaagtgct	2220
Db	2161	AGAGTTGTGTGAAGACCAAGCGTGGGTTCATGTGAACGTGGCATGTGACAGGAGGTGT	2220
Qy	2221	ggaccccccaattgaaacgttgaagtaacttaagaactccaagaagtgatgtgaatgccat	2280
Db	2221	GGACCCCCCAATTGAAGCGTGAATACCTTACAGCCTTCAAGACAGTGTATGCCAT	2280
Qy	2281	tgcccaagagttctcaactatgtgtctctagtctccgcgggttctgtatgtctgttgaag	2340

```
D 2281 TGCACAGATTCTCAGCTGATGTGTCCTAGTCTCCGCGGTTTGATGATCTGTTGAAG 2340
Q 2341 acattgtctctctctgtgtgtgtactctgtctccgcgcagatgttttggccaattaccag 2400
D 2341 ACATGTGCTCTCTGTGGGGGTACTCTGTCAACCCGCAATATTTTGGCCACTTACCG 2400
Q 2401 gaagctgtgaacctggcagaggggaggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2460
D 2401 GCAGCTGATGACCTTCGCGAGGGGGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2460
Q 2461 gaccgcacatgtgtatgtacctgtgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2520
D 2461 GACCGCCATCTGTGATGCTCTGTAAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2520
Q 2521 gacctgtgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2580
D 2521 GCGCTTGATGTGAGGAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2580
Q 2581 gaaatcatcgtgacatcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 2640
D 2581 GAAATCATCTGAGATCCAGAGCAAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2640
Q 2641 gggccggtccctgtcgaaggggcccgaagcgaagcgaagcgaagcgaagcgaagcgaagc 2700
D 2641 GGGCGGTCCTGTGCGAGGGGGCCCAAGCAGGTGAGACCGAAGCCGAATGTGAACGCC 2700
Q 2701 atgtcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2760
D 2701 ATGCGCTTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2760
Q 2761 cccaaagccgagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 2820
D 2761 CCCAAGCCGCGAGAGAGCCCATGAGAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGT 2820
Q 2821 ccccttgggtcctcacatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2880
D 2821 CCGTGTGGGCTTCAACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2880
Q 2881 aattt 2885
D 2881 AATTT 2885

RESULT 2
V73805/c
ID V73805 standard; DNA: 32207 BP.
AC V73805;
DE 25-FEB-1999 (first entry)
DE KSHV LUR DNA (nucleotides 105,301-137,507).
KW Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;
KW dilydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;
KW glycoprotein; kaposin; cyclin D; immediate early protein; IIP; OX-2;
KW v-ah; G-protein coupled receptor; FGARAT; ds.
OS Kaposi's sarcoma-associated herpesvirus.
PN US849564-A.
PD 15-DEC-1998.
PR 29-NOV-1996; 770379.
PR 29-NOV-1996; US-770379.
PA (UYCO ) UNIV COLUMBIA NEW YORK.
PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
DR WPI: 99-069741/06.
PT Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
PT dihydrofolate reductase and is useful for treatment, prophylaxis
PT or diagnosis of Kaposi's sarcoma
PS disclosure: Column 155-182; 109pp; English.
CC This sequence is a fragment of the Kaposi's sarcoma-associated
CC herpesvirus (KSHV) LUR (long unique region). This fragment contains
CC coding regions for ORF65 which encodes capsid protein IV, ORF66, ORF67
CC which encodes tegument protein IV, ORF68 which encodes a glycoprotein,
CC ORF69, K12 which encodes kaposin, K13, ORF72 which encodes cyclin D,
CC ORF73 which encodes immediate early protein (IIP), K14 which encodes
CC OX-2 (v-ah), ORF74 which encodes G-protein coupled receptor, ORF75
```

```
CC which encodes tegument protein/FGARAT. K15. KSHV is a new human
CC Herpesvirus (HHV8) believed to cause Kaposi's sarcoma (KS) which is the
CC most common form of neoplasia occurring in persons with acquired immune
CC deficiency syndrome (AIDS). The DHFR protein is useful for vaccination,
CC prophylaxis, diagnosis and treatment of a subject with Kaposi's sarcoma
CC and for detecting expression of a DNA virus associated with Kaposi's
CC sarcoma in a cell.
SQ Sequence 32207 BP: 7229 A; 9156 C; 8713 G; 7109 T;

Query Match 2.9%; Score 84.6; DB 1; Length 32207;
Best Local Similarity 48.8%; Pred. No. 4.6e-09;
Matches 293; Conservative 0; Mismatches 299; Indels 9; Gaps 2;

Q 770 ctgcagcatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 829
D 20394 CAGCAGCAGAGGACCCACAGCAGAGGAGGCGCACAGCAGAGGAGCAGCAGAGGAGC 20335
Q 830 ctccacgggacagtcctccatagtgacggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 889
D 20334 CAGCAGCGGGAGCCCGCAGCAGCGGGAGCCACAGCAGCGGGAGCCACAGCAGCGG 20275
Q 890 ggcgaagctcccgccgcatcgcgcctgtgagccgactcaagctccacgcgtgcgcag--- 947
D 20274 CAGCAGCGGGAGCCCGCAGCAGCGGGAGCCACAGCAGCGGGAGCCACAGCAGCGG 20215
Q 947 agtcccaagccctcgaagcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1006
D 20214 CAGCAGCAGAGATGACAGCAGCAGAGATGACAGCAGCAGAGATGACAGCAGCAG 20155
Q 1007 cagaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1066
D 20154 CAGCAGCAGAGATGACAGCAGCAGAGATGACAGCAGCAGAGATGACAGCAGCAG 20095
Q 1067 cagccaccaccaccaccctgtgagcagcagcagcagcagcagcagcagcagcagcagc 1126
D 20094 CAGCAGCAGAGATGACAGCAGCAGAGATGACAGCAGCAGAGATGACAGCAGCAG 20035
Q 1127 ctgggggagggagccctgacatgcctcccgaggaggtccacagagatgtgagcagacag 1186
D 20034 CAGCAGCAGAGATGAC-----AGCAGCAGAGATGACAGCAGCAGAGATGACAGCAG 19981
Q 1187 gaagacctgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1246
D 19980 GATTGACAGCAGCAGAGATGACAGCAGCAGAGATGACAGCAGCAGAGATGACAGCAG 19921
Q 1247 gttaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1306
D 19920 GATTGACAGCAGCAGCAGAGATGACAGCAGCAGAGATGACAGCAGCAGAGATGACAG 19861
Q 1307 gctgatacaaaaaactgtctcagatgtcccaaccgctgtcaaccttgcaggtgtacca 1366
D 19860 GAGCAGCAGAGATGACAGCAGCAGAGATGACAGCAGCAGCAGAGATGACAGCAGAG 19801
Q 1367 g 1367
D 19800 G 19800

RESULT 3
V19941/c
ID V19941 standard; DNA: 137507 BP.
AC V19941;
DE 03-AUG-1998 (first entry)
DE KSHV long unique coding region 8; macrophage inflammatory protein II;
KW KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
KW interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
KW complement-binding protein; glycoprotein; capsid protein IV; infection;
KW lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;
KW HIV immune status; anti-inflammatory agent; therapy; ds.
OS Kaposi's sarcoma-associated herpes virus.
FH key Location/Qualifiers
```

FT	CDS	1142. .2794	/*tag= a
FT		/product= complement-binding protein	
FT	CDS	8699. .11236	/*tag= b
FT		/product= glycoprotein B	
FT	CDS	complement (17261. .17875)	
FT		/*tag= c	
FT		/product= interleukin 6	
FT	CDS	/complement (21548. .21832)	
FT		/*tag= d	
FT		/product= macrophage inflammatory protein II	
FT	CDS	complement (27137. .27424)	
FT		/*tag= e	
FT		/product= interferon regulatory factor 1	
FT	CDS	2861. .29741	/*tag= f
FT		/product= protein T1.1	
FT	CDS	complement (58976. .60175)	
FT		/*tag= g	
FT		/product= glycoprotein M	
FT	CDS	complement (69412. .69915)	
FT		/*tag= h	
FT		/product= glycoprotein L	
FT	CDS	complement (88410. .88910)	
FT		/*tag= i	
FT		/product= interferon regulatory factor 2	
FT	CDS	89600. .90541	/*tag= j
FT		/product= interferon regulatory factor 3	
FT	CDS	90173. .90643	/*tag= k
FT		/product= glycoprotein X	
FT	CDS	complement (93636. .94127)	
FT		/*tag= l	
FT		/product= interferon regulatory factor 4	
FT	CDS	complement (111931. .112443)	
FT		/*tag= m	
FT		/product= capsid protein IV	
FT	CDS	complement (123808. .127296)	
FT		/*tag= n	
FT		/product= immediate early protein	
PD		MO9804576-A1.	
PD		05-FEB-1998.	
PF		22-JUL-1997; U13346.	
PR		29-NOV-1996; PR-757669.	
PR		25-JUL-1996; US-686243.	
PR		25-JUL-1996; US-686349.	
PR		25-JUL-1996; US-686350.	
PR		25-JUL-1996; US-687253.	
PR		25-JUL-1996; US-688814.	
PR		05-SEP-1996; US-708678.	
PR		10-OCT-1996; US-728323.	
PR		13-NOV-1996; US-747887.	
PR		13-NOV-1996; US-748640.	
PA		(UTCO ) UNIV COLUMBIA NEW YORK.	
PI		Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;	
PI		WPI; 98-130615/12.	
PT		New nucleic acid encoding Kaposi's sarcoma associated herpes virus	
PT		proteins - useful for e.g. detecting levels of HHV8 in, and	
PT		preparation of vaccines for treatment of, HIV patients	
PS		Example 2; Page 135-203; 230pp; English.	
CC		This sequence represents the long unique region and terminal repeat of	
CC		the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known	
CC		as human herpes virus 8 (HHV8). This sequence contains the DNAs of the	
CC		invention which encode KSHV polypeptides selected from: (a) viral	
CC		microphage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);	
CC		(c) viral INF 1; (d) complement-binding protein; glycoproteins B, M or L;	
CC		(d) capsid protein IV encoded by ORF6; and (e) immediate early protein	
CC		encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded	
CC		by it, and antibodies (Ab) specific for the proteins are useful for	
CC		detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body	
CC		fluids or tissue samples. HHV8 infections can be treated with antisease	

or triplex forming molecules or agents that bind specifically to the protein. Ab may be used for prophylaxis or treatment of HHV8 infection, while the protein can be used in protective vaccines. Ab may also be used to differentiate between lymphomas, and HHV8 may be implicated in many other lymphoproliferative diseases such as lymphomas, leukemias, splenomegaly and mycosis fungoides. Cells and animals containing the nucleic acid are useful for drug screening. HHV8-derived peptides can be used as targets for antiviral drugs, e.g., diphthololate reductase gene can be inhibited with methotrexate. These can also be used to determine the immune status of a patient infected with HIV. HHV8 derived protein viral MIP III may be used as an anti-inflammatory agent for, e.g., treating rheumatoid arthritis. This sequence is stated as containing 81 open reading frames.

Sequence 137507 BP: 32579 A; 37795 C; 35758 G; 31375 T;

Query Match	2.9%;	Score 84.6;	DB 1;	length 137507;
Best Local Similarity	48.8%;	Pred. No. 7e-09;		
Matches 293;	Conservative	0;	Mismatches 299;	Indels 9;
				Gaps 2;

QY	770	ctgaagatgtgctggttcttgagacaagcccgagagaagaacccctcatgtctgtgcca	829
Db	125664	CAGCAGCAGGAGCCACACACACACAGACCCACACAGCAGGAGGCCACACAGCCGGAGCCA	125635
QY	830	ctccaagcgagcttccccaactgtacaggtgtgaagctgtggccaaccaatctcgagagcgtta	889
Db	125634	CAGCAGCGGGAGGCCCCACAGACCGGGAGCCACACAGCAGCGGGAGGCCACAGCCGGAGCCA	125575
QY	880	ggcaagctcccgcgcatctggtcccttgagccgcaactagtctctcaaccgtctgcgcag---	947
Db	125574	CAGCAGCGGGAGCCACACACAGCCGGAGCCACACAGCAGCGGGAGCCACACAGCCGGAGCCA	125515
QY	947	agtccecaagcgcctgcagcaagctgtcatgtcaacaacagcaccaagcattctcttgagaag	1006
Db	125514	CAGCAGCAGGATGTAGACACACAGAGATGTAGACAGCAGCAGCAGATGAGCAGCAGCAGATGAG	125455
QY	1007	cagaagcagcagcagctacacagcttggccaagatcctcaaccaagaaggagactgtgccag	1066
Db	125454	CAGCAGCAGGATGTAGACACAGCAGAGATGTAGACAGCAGCAGATGAGCAGCAGCAGATGAG	125395
QY	1067	cagcccaaccaaccaccccttgaggagacagagaggagactgtacagcagcagcggtcttg	1126
Db	125394	CAGCAGCAGGATGTAGACACACAGAGATGTAGACAGCAGCAGATGAGCAGCAGCAGATGAG	125335
QY	1127	ctgggggagaggagcccttgaccatgtcccccggagggtctccacagagatgtgagacaacag	1186
Db	125334	CAGCAGCAGGATGTAGC-----AGCAGCAGAGATGTAGCAGCAGCAGATGTAGCAGCAGCAG	125281
QY	1187	gaagaccttgagagagagagagagaaagagatgtggagagagagagatgtcatccag	1246
Db	125280	GATGAGCAGCAGCAGGATGTAGACACACAGAGATGTAGACAGCAGCAGATGAGCAGCAGCAG	125221
QY	1247	gttaagacagagagggcgagagatgtgtgcctgagagagggccgaacttgagagagcttgt	1306
Db	125220	GATGAGCAGCAGCAGGATGTAGACAGCAGCAGAGGATGTAGCAGCAGCAGCAGATGTAGCAG	125161
QY	1307	gctgatatcaaaaacagcttctcagatgtgcccaaccgcttgacaacttgcaggtgtatcca	1366
Db	125160	GAGCAGCAGGATGTAGCAGCAGCAGCAGAGATGTAGCAGCAGCAGGATGTAGCAGCAGCAG	125101
QY	1367	g 1367	
Db	125100	G 125100	
RESULT	4		
ID	V55831		
AC	V55831		
DT	18-NOV-1998	(first entry)	
DE	Nucleotide sequence of the stabilising sequence-encoding insert.		
FW	Fusion protein; stabilising polypeptide; proteolytic degradation;		
KW	resistance; half-life; autoimmune disease; inflammation; nitro drug;		



```
Qy 1063 cagcagccaccaccacccttgagagacagagagctgacgagcagcagagat 1122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 955 GGAGCAGAGAGAGAGGCGAGAGGCGCAGAGAGAGAGAGAGAGAGAGAGAGG 1014
Qy 1123 ctgtctggaggagagagccctgaccatgcccgcggagaggtccacagagtgtagagcac 1182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1015 GCAGGAGCAGAGAGAGGCGCAGAGCAGAGAGAGAGAGAGAGAGAGAGAGCA 1074
Qy 1183 acagggaagacctgagagagagagagcagagagatggtgagagagagagatgcat 1242
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1075 GGAGGGCGCAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGAGAGAGAGAGGCA 1134
Qy 1243 ccagggttaagagcagagagagcagagagtggtgtctgagagagagagccgacttgagagagc 1301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1135 GGAGCAGAGAGAGGCGCAGAGAGAGAGGCGCAGAGAGAGAGAGAGAGAGAGC 1193

RESULT 6
ID 051731 standard; DNA; 10596 BP.
AC 051731:
DE 31-MAY-1994 (first entry)
KW Plasmid pCISEBON for subcloning huHGF variants.
KW Hepatocyte growth factor; HGF; variant; mutain; in vitro mutagenesis;
KW proteolysis resistant; liver; malignancy; CMV-driven;
KW Cytomegalovirus; episomal expression plasmid; ss.
OS Synthetic.
FH Key
FT enhancer
    /tag- a
    /note- "CMV enhancer/promoter"
    /tag- b
    /label- Sp6-promoter
    /tag- c
    /note- "Sp6 RNA start"
    /tag- d
    /function- cloning_linker
    /tag- e
    /note- "SV40 poly A"
    /tag- f
    /function- SV40-origin
    /tag- g
    /label- EBNA-1
    /tag- h
    /function- oriP
    /tag- i
    /note- "family of repeats"
    /tag- j
    /note- "quad region"
    /tag- k
    /label- HSV_TK_terminator_3'-end
    /tag- l
    /phenotype- neomycin_resistance
    /note- "Tr5 neomycin phosphotransferase gene"
    /tag- m
    /label- TK_promoter
    /tag- n
    /function- M13_ori
    /tag- o

misc_feature
polya_signal
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
repeat_region
misc_structure
terminator
cds
promoter
misc_feature
misc_rna
```

```
FT 109323541-A. /label- delta_2a
PN 25-NOV-1993.
PF 17-MAY-1993; U04648.
PR 18-MAY-1992; US-884811.
PR 18-MAY-1992; US-885971.
PA (GETH ) GENENTECH INC.
PI Godowski PJ, Lokker NA, Mark MR;
DR WPI: 93-386573/48.
PT Hepatocyte growth factor variants - are resistant to proteolytic
PT cleavage into its two-chain form, used to treat malignancies
PT associated with HGF receptor
PS Example 1; Fig 6; 87bp; English.
CC Plasmid pCISEBON (a pRK5 derivative) is an episomal CMV driven
CC expression plasmid. HuHGF variants with enhanced receptor binding
CC activity were produced by site-directed mutagenesis. Stable
CC populations of preferred HGF variants were obtained by transfecting
CC human embryonic kidney 293 cells and then these were subcloned in
CC pCISEBON. See R52940-R52949 for examples of pref. HGF variants.
SQ Sequence 10596 BP; 2625 A; 2571 C; 3024 G; 2376 T;

Query Match 2.7%; Score 79; DB 1; Length 10596;
Best Local Similarity 51.3%; Pred. NO. 5.3e-08;
Matches 184; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

Qy 943 gcagagtcgccagcctgtgacagctggtcatgcaacaagcagcagcagcttcgga 1002
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2369 GCAGAGAGGCGAGAGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2428
Qy 1003 gaagcagaagcagcagcagctacagctgtgcaagatcctccaaagacagggagctgccc 1062
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2429 GGGCGCAGAGAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGAGAGAGAGAGAG 2488
Qy 1063 cagcagccaccaccacccttgagagagagagagagagagctgacagagcagagat 1122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2489 GCAGCAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGAGAGAGAGAGAGAGGG 2548
Qy 1123 ctgtctggaggagagagccctgaccatgcccgcggagaggtccacagagatgtagagcac 1182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2549 GCAGAGAGCAGAGAGAGGCGCGAGAGCAGAGAGGCGCAGAGAGAGAGAGAGCA 2608
Qy 1183 acagggaagacctgagagagagagcagagagatggtgagagagagagagatgcat 1242
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2609 GGAGGGCGCAGAGAGAGAGAGGCGCAGAGAGGCGCAGAGAGAGAGAGAGGCA 2668
Qy 1243 ccagggttaagagcagagagagcagagtggtgtctgagagagagagccgacttgagagagc 1301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2669 GGAGCAGAGAGGCGCAGAGAGAGGCGCAGAGAGAGAGAGAGAGAGAGAGAGC 2727

RESULT 7
ID T40348 standard; DNA; 10596 BP.
AC T40348;
DE 09-DEC-1996 (first entry)
DE Plasmid pCISEBON for expression of hepatocyte growth factor.
KW Human; hepatocyte growth factor; HGF; huHGF; serum; proteolytic cleavage;
KW pro-hormone; beta subunit; alpha subunit; kringle domain; prothrombin;
KW plasminogen; catalytic domain; serine protease; HGF variant;
KW HGF receptor; malignancy; chronic HGF receptor activation; ss.
OS Synthetic.
PN US5547856-A.
PD 20-AUG-1996.
PF 18-MAY-1992; 884811.
PR 18-MAY-1992; US-885971.
PR 18-MAY-1992; US-884811.
PR 13-JUL-1993; US-087783.
PA (GETH ) GENENTECH INC.
PI Godowski PJ, Lokker NA, Mark MR;
DR WPI: 96-392634/39.
PT New hepatocyte growth factor variants - are resistant to in vivo
PT proteolytic cleavage into a 2-chain form, useful as HGF antagonists
```







AC V87399;  
 DE 27-APR-1999 (first entry)  
 KM Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 KM tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 KM chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
 KM receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
 OS Homo sapiens.  
 PN M09845435-A2.  
 PD 15-OCT-1998.  
 PF 10-APR-1998; 006954.  
 PR 10-APR-1997; US-835913.  
 PA (GENY ) GENETICS INST INC.  
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
 PI Racle LA, Spaulding V, Treacy M;  
 DR WPI: 99-07076/06.  
 PT New polynucleotides encoding human secreted proteins - derived from  
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon cDNA libraries  
 PS Claim 1: Page 551: 633pp. English.  
 CC This sequence represents an expressed sequence tag (EST), and is a  
 CC polynucleotide of the invention. The polynucleotides of the invention are  
 CC all secreted EST sequences isolated from a variety of human tissue  
 CC sources. The EST sequences and proteins encoded by them are predicted to  
 CC have useful biological activities which would make them suitable for  
 CC treating, preventing or ameliorating medical conditions in humans and  
 CC animals, although no supporting data is given. Suggested activities  
 CC include nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The EST sequences are also stated to be useful for gene  
 CC therapy.  
 SQ Sequence 424 BP; 89 A; 140 C; 109 G; 86 T;

Query Match 2.3%; Score 66.4; DB 1; Length 424;  
 Best Local Similarity 61.0%; Pred. No. 1.1e-05;

Matches 128; Conservative 0; Mismatches 76; Indels 6; Gaps 1;

OY 2281 tggccagaggttcctaccatggtgctctgctcgcgcgggttgagctgtgaagg 2340  
 DB 15 tggccttagagtttaacccagaaactgctgctgctcagcttgatgctt-----gc 68  
 OY 2341 aacatgtctcctctggtggtgctactctgtcacgcgcagatgtttgaccattgaccag 2400  
 DB 69 acgggggggattccgctgggggggctgcacagtgacactgaggggttaagccacactaccca 128  
 OY 2401 gcaagtgtatgacctggcaggggggagcgggtgtgtctgcccctggaggagggcactgactt 2460  
 DB 129 cctgctgtaggggcttgccagtcagtcgacgtttatcctttatcctagaggggtgctatfaacct 188  
 OY 2461 gaccgcacatctgtatgctcctctgaagcttg 2490  
 DB 189 gacatccatctcagagctccatgctgctg 218

RESULT 12  
 O84832

ID O84832 standard; DNA; 234 BP.

AC O84832;

DE 25-SEP-1995 (first entry)

KM Spinocerebellar ataxia type 1 CAG repeat region patient #2.

KM Spinocerebellar ataxia type 1; SCA 1; presymptomatic diagnosis;

KM CAG repeat region; patient #2; ss.

OS Homo sapiens.

PN W09501437-A.

PD 12-JAN-1995.

PF 29-JUN-1994; 007336.

PR 29-JUN-1993; US-084365.

PR 28-JUN-1994; US-267803.

PA (MINU ) UNIV MINNESOTA.

PI Chung M, Orr HT, Zoghbi HY;  
 DR WPI: 95-061001/08.  
 PT New autosomal dominant spinocerebellar ataxia type 1 nucleic acid  
 PT - used to develop prods. for detection or presymptomatic  
 PT diagnosis of a SCA1 disorder  
 PS Disclosure; Fig 2; 111pp. English.  
 CC 084831-084835 show the CAG repeat regions of five individuals  
 CC affected with spinocerebellar ataxia type 1 (SCA 1). It is within  
 CC CAG repeat region (084804) that the mutations responsible for  
 CC SCA 1 occur. The full nucleic acid (084793) and its protein product  
 CC (R7111) can be used to develop products, for the presymptomatic  
 CC detection of a SCA 1 disorder.  
 SQ Sequence 234 BP; 73 A; 80 C; 77 G; 4 T;

Query Match 2.3%; Score 65; DB 1; Length 234;  
 Best Local Similarity 55.6%; Pred. No. 1.8e-05;

Matches 125; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

OY 935 ccgctgcccagagatcccaagccctgcagcagctgtgtatgaaacaagccagcg 994  
 DB 4 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 63  
 OY 995 ttccctggagaagcagaagcagcagcagctacagctgtggcaatcctcaacaagcaggg 1054  
 DB 64 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 123  
 OY 1055 gagctgcccagcagcccaaccacccctgaggagacagagagagctgcagcag 1114  
 DB 124 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 183  
 OY 1115 caggaggtctgtctggtgggagggagccctgacatgccccggag 1159  
 DB 184 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 228

RESULT 13  
 O23646

ID O23646 standard; DNA; 1032 BP.

AC O23646;

DE 18-SEP-1992 (first entry)

KM Acetylpolylamine amidohydrolase gene.

KM Tumour; polylamine; urea; PAH; ss.

OS Mycoplasma bullata FERM BP-1845

PN J04071489-A.

PD 06-MAR-1992.

PF 12-JUL-1990; 184979.

PR 12-JUL-1990; JP-184979.

PA (KYOW ) KYOWA HAKKO KOGYO KK.

DR WPI: 92-128229/16.

P-PSDB: R22463.

PT Acetyl poly:amine amido hydrolase prepn. for diagnosis - by

PT culturing microbe of Escherichia and Mycoplasma and recombinant

PT deoxyribonucleic acid, for tumour marker

PS Claim 5; Page 2; 9pp; Japanese.

CC The protein produced from this sequence can be used for determining

CC polyamine in urea which is a tumour marker.

SQ Sequence 1032 BP; 178 A; 344 C; 320 G; 190 T;

Query Match 2.2%; Score 63.8; DB 1; Length 1032;  
 Best Local Similarity 52.4%; Pred. No. 5.1e-05;

Matches 165; Conservative 0; Mismatches 147; Indels 3; Gaps 1;

OY 1886 gaagagagatcaagaatgattgcatcatccggcccccagaccagcccgagaa 1945  
 DB 433 GCGCGCGCCACAGAGCCGCGCTTCTGCTTGCCGCCCGCGGGGACCAAGCCGCGCAATC 492  
 OY 1946 tccacagcattggaattctgtctctcaactctgttagcatcacccgcaaaactctacag 2005  
 DB 493 GACATGTTGGGGGCTAGTGCCTTCAATCAACAAGCGCGCGTGCAGCCAGCGGCTGTC 552  
 OY 2006 cagaagtgaagctggcgaaggtcctcatctcgtgactgtggaattcaccattggaattgac 2065



OY 1295 gagag 1300  
||  
DB 1004 GACAG 1009

Search completed: May 20, 2000, 00:36:50  
Job time: 11862 sec

**THIS PAGE BLANK (USPTO)**

US-08-232-463-14

Page 2





TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32207 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-770-379-20

Query Match 2.7%; Score 58.8; DB 3; Length 32207;  
Best Local Similarity 52.4%; Pred. No. 6.8e-06;  
Matches 129; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 978 gagagcgttcattacagacacggcgagcggttgagagcgccgagcagcgtgagctgcagc 1037  
DB 19813 GCAGCGACGACGAGGATGACGACGACGACGATGACGACGACGACGACGATGACGACG 19754  
QY 1038 gcagagcgttcattacagacacggcgagcggttgagagcgttcacacagatccctcagagca 1097  
DB 19753 GCAGCGACGATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 19694  
QY 1098 gcagagcgttcattacagacacggcgagcggttgagagcgttcacacagatccctcagagca 1157  
DB 19693 GTTAGAGGACGACGACGACGACGATGACGATGACGACGACGATGACGACGACGACGAC 19634  
QY 1158 gagatcccgagagagagatgagacagattgtgagagagagagagagatttaagaagcaatg 1217  
DB 19633 GCGATTAGAGGACGACGACGACGACGATGACGACGACGACGACGATGACGACGACGACG 19574  
QY 1218 ggaaga 1223  
DB 19573 GCAGGA 19568

RESULT 4  
US-09-130-114-1/c  
; Sequence 1, Application US/09130114  
; Patent No. 5976807  
; GENERAL INFORMATION:  
; APPLICANT: Horlick, Robert A.  
; APPLICANT: Dama, Basam B.  
; APPLICANT: Robbins, Alan K.  
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes  
; TITLE OF INVENTION: From Multiple Transfected Episomes  
; FILE REFERENCE: 0867/1D903US1  
; CURRENT APPLICATION NUMBER: US/09/130,114  
; CURRENT FILING DATE: 1998-08-06  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 5452  
; TYPE: DNA  
; ORGANISM: YEENA  
US-09-130-114-1

Query Match 2.6%; Score 57.2; DB 4; Length 5452;  
Best Local Similarity 50.7%; Pred. No. 7e-06;  
Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 976 cggagagcgttcattacagacacggcgagcggttgagagcgccgagcagcgtgagctgcag 1035  
DB 2138 CGGAGGCGGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2079  
QY 1036 cggagagcgttcattacagacacggcgagcggttgagagcgccgagcagcgtgagctgcag 1095  
DB 2078 CAGGAGGCGGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2019  
QY 1096 cagcagagatgagcgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1155

DB 2018 GCGACGAGGCGGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1959  
QY 1156 gagagatcccgagagagatgagacagattgtgagagcgagagagagagagagagagagca 1215  
DB 1958 GAGGGGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1899  
QY 1216 tggagagagagcgtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1245  
DB 1898 CAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1869

# RESULT 5

US-07-884-811-15  
; Sequence 15, Application US/07884811  
; Patent No. 5316921

; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: patin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/884,811  
; FILING DATE: 19920518  
; CLASSIFICATION: 330  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dregger, Ginger R.  
; REGISTRATION NUMBER: 33,055  
; REFERENCE/DOCKET NUMBER: 755.1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-3216  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10596 bases  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-07-884-811-15

Query Match 2.6%; Score 57.2; DB 1; Length 10596;  
Best Local Similarity 50.7%; Pred. No. 1e-05;  
Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 976 cggagagcgttcattacagacacggcgagcggttgagagcgccgagcagcgtgagctgcag 1035  
DB 2247 CGGAGGCGGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2306  
QY 1036 cggagagcgttcattacagacacggcgagcggttgagagcgccgagcagcgtgagctgcag 1095  
DB 2307 CAGAGGCGGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1095  
QY 1096 cagcagagatgagcgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1155  
DB 2367 GCGACGAGGCGGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1155  
QY 1156 gagagatcccgagagagatgagacagattgtgagagagagagagagagagagagagagag 1155

Page

Db 2487 CAGGAGCAGGAGGCGCAGGAGGGCAG 2516

RESULT 7  
US-08-087-783A-15  
; Sequence 15, Application US/08087783A

```

1  APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
2  TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS
3  NUMBER OF SEQUENCES: 22
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE: Genentech, Inc.
6  STREET: 460 Point San Bruno Blvd
7  CITY: South San Francisco
8  STATE: California
9  COUNTRY: USA
10 ZIP: 94080
11
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOOS
16 SOFTWARE: Winpatlin (Genentech)
17
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/087,783A
20 FILING DATE: 13-Jul-1993
21 CLASSIFICATION: 435
22 PRIOR APPLICATION DATA:
23 PRIOR APPLICATION NUMBER: 07/884811
24 FILING DATE: 18-MAY-92
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: 07/885971
27 FILING DATE: 18-MAY-92
28 ATTORNEY/AGENT INFORMATION:
29 NAME: Marschang, Diane L.
30 REGISTRATION NUMBER: 35,600
31 REFERENCE/DOCKET NUMBER: P0755779P1
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: 415/225-5416
34 TELEFAX: 415/952-9881
35 TELEX: 910/371-7168
36 INFORMATION FOR SEQ ID NO: 15:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 10596 base pairs
39 TYPE: Nucleic Acid
40 STRANDEDNESS: Single
41 TOPOLOGY: Linear
42
43 US-08-087-783A-15

```

Query Match	2.6%	Score 57.2	DB 1	Length 10536
Best Local Similarity	50.7%	Pred. No. 1e-05		
Matches 137	Conservative	0	Mismatches 133	Indels 0
				Gaps 0
QY 976	cggagagctgttcattatpacacgacccggygaagcgygtctgtgcagagagcgcgagcagctgtgagcttcag	1035		
Db 2247	CGGGGGGGGCGAGGACAGAGAGGGCGAGGACAGGACAGGAGGAGGGCAGGACAGAGAGGAGGGG	2306		
QY 1036	cggcagagagcttcattatcagaagaacgagcttgcgtagtgagatgagatccaataatcttcagagag	1095		
Db 2307	CAGGAGGGGCGAGAGGGGCGAGGAGGGGCGAGACCGAGGAGAGGGCAGGACGAGGAGGAGG	2366		
QY 1096	cagcagagagatltgagcgcgcaagaagagaaagaatltgcccagaagcgacagcagagaaat	1155		
Db 2367	GGCGACAGGGGGCGAGAGGGGCGACGAGCAGAGGAGGAGGGGGCGAGCAGGAGGAGGGGCGAG	2426		
QY 1156	gagagatacccggaagagagatltgaacaattttagagagaggaagaagttaagaagaa	1215		
Db 2427	GAGGGCGAGGACGAGGAGGAGGGGCGAGAGGGGCGCAGGAGGGGCGAGGACGAGGAGGAGGGG	2486		
QY 1216	tggygaagaagagacttgyggctccaagagaaacag	1245		
Db 2487	CAGGAGCAGGAGGAGGGGCGAGGAGGGGCGAG	2516		

```

RESULT      8
US-08-194-088B-15
; Sequence 15, Application US/08194088B
; Patent No. 5580963
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,088B
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallegos, R. Thomas
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: 755D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-2614
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-194-088B-15

```

```

Query Match      2.6%; Score 57.2; DB 1; Length 10596;
Best Local Similarity 50.7%; Pred. No. 1e-05;
Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 976 cggagctgttcattgacagaccgagcgctgagcagagcgagcgctgagcgag 1035
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2247 CGGAGGGCGCAGAGCAGAGGGGCGAGGCGAGGAGGCGAGGAGGAGGAGG 2306
QY 1036 cggagagcttctcattgacagagcgctgagcagagcgagcgctgagcgag 1095
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2307 CAGGAGGGCGCAGAGGCGAGGCGAGGCGAGGAGGCGAGGAGGCGAGGAGG 2366
QY 1096 cagcagagagtgagcgagcagagagagagagagagagagagagagagag 1155
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2367 GGGCAGAGAGGGCGAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2426
QY 1156 gagagataccgagagagagagagagagagagagagagagagagagagag 1215
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2427 GAGGGGCGAGAGCAGAGGAGGCGAGGAGGCGAGGAGGCGAGGAGGAGGAGG 2486
QY 1216 tgggagagagagagagagagagagagagagagagagagagagagagag 1245
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2487 CAGGAGCAGAGAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2516

```

RESULT 9  
US-08-194-087-15

```

; Sequence 15, Application US/08194087
; Patent No. 5879910
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,087
; FILING DATE: 18-MAY-1992
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dieger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-194-087-15

```

```

Query Match      2.6%; Score 57.2; DB 3; Length 10596;
Best Local Similarity 50.7%; Pred. No. 1e-05;
Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 976 cggagctgttcattgacagaccgagcgctgagcagagcgagcgctgagcgag 1035
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2247 CGGAGGGCGCAGAGCAGAGGGGCGAGGCGAGGAGGCGAGGAGGAGGAGG 2306
QY 1036 cggagagcttctcattgacagagcgctgagcagagcgagcgctgagcgag 1095
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2307 CAGGAGGGCGCAGAGGCGAGGCGAGGCGAGGAGGCGAGGAGGCGAGGAGG 2366
QY 1096 cagcagagagtgagcgagcagagagagagagagagagagagagagagag 1155
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2367 GGGCAGAGAGGGCGAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2426
QY 1156 gagagataccgagagagagagagagagagagagagagagagagagagag 1215
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2427 GAGGGGCGAGAGCAGAGGAGGCGAGGAGGCGAGGAGGCGAGGAGGAGGAGG 2486
QY 1216 tgggagagagagagagagagagagagagagagagagagagagagagag 1245
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2487 CAGGAGCAGAGAGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2516

```

RESULT 10  
PCT-US93-04648-15  
; Sequence 15, Application PC/TUS9304648  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie  
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS



RESULT 14  
 US-08-056-200-93  
 ; Sequence 93, Application US/08056200  
 ; Patent No. 5616500  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Steinerl, Peter M.  
 ; APPLICANT: Lee, Seung-Chul  
 ; APPLICANT: Kim, In-Gyu  
 ; APPLICANT: Chung, Soo-Il  
 ; APPLICANT: Park, Sang-Chul  
 ; TITLE OF INVENTION: Trichothyalin and Transglutaminase-3 and  
 ; TITLE OF INVENTION: Methods of using Same  
 ; NUMBER OF SEQUENCES: 117  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear  
 ; STREET: 620 Newport Center Drive, Sixteenth Floor  
 ; CITY: Newport Beach  
 ; STATE: CA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 92660  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/056,200  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507..1644  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1645..2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512..8070  
US-08-056-200-93

Query Match 2.2%; Score 48.2; DB 1; Length 9551;  
Best Local Similarity 48.4%; Pred. No. 0.0027;  
Matches 134; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 973 ggcgggagctgttcatacagaccggagcgctgagcagcgcgagcgtgagctg 1032  
DB 4888 ggcgggagctgttcatacagaccggagcgctgagcagcgcgagcgtgagctg 1032  
QY 1033 ggcgggagctgttcatacagaccggagcgctgagcagcgcgagcgtgagctg 1092  
DB 4948 ggcgggagctgttcatacagaccggagcgctgagcagcgcgagcgtgagctg 1092  
QY 1093 ggcgggagctgttcatacagaccggagcgctgagcagcgcgagcgtgagctg 1152  
DB 5008 ggcgggagctgttcatacagaccggagcgctgagcagcgcgagcgtgagctg 1152  
QY 1153 aatgagagatccggaagagatggaacagattgttagagagagagagatttaagag 1212  
DB 5068 aatgagagatccggaagagatggaacagattgttagagagagagagatttaagag 1212  
QY 1213 caatgggaagaagactggtggtcctcaaggaacagctac 1249  
DB 5128 gaacagacactgctgcagagagagagagagagagagagagagagagagagagag 5164

RESURF 15  
US-08-800-644-93  
Sequence 93, Application US/08800644  
Patent No. 5958752  
GENERAL INFORMATION:  
APPLICANT: Steinert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-gyu  
APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knudde, Martens, Olsson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor

CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/800,644  
FILING DATE: 14-FEB-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/056,200  
FILING DATE: 30-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507..1644  
FEATURE:  
NAME/KEY: intron  
LOCATION: 1645..2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512..8070  
US-08-800-644-93

Query Match 2.2%; Score 48.2; DB 4; Length 9551;  
Best Local Similarity 48.4%; Pred. No. 0.0027;  
Matches 134; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 973 ggcgggagctgttcatacagaccggagcgctgagcagcgcgagcgtgagctg 1032  
DB 4888 ggcgggagctgttcatacagaccggagcgctgagcagcgcgagcgtgagctg 1032  
QY 1033 ggcgggagctgttcatacagaccggagcgctgagcagcgcgagcgtgagctg 1092  
DB 4948 ggcgggagctgttcatacagaccggagcgctgagcagcgcgagcgtgagctg 1092  
QY 1093 ggcgggagctgttcatacagaccggagcgctgagcagcgcgagcgtgagctg 1152  
DB 5008 ggcgggagctgttcatacagaccggagcgctgagcagcgcgagcgtgagctg 1152  
QY 1153 aatgagagatccggaagagatggaacagattgttagagagagagagatttaagag 1212  
DB 5068 aatgagagatccggaagagatggaacagattgttagagagagagagatttaagag 1212  
QY 1213 caatgggaagaagactggtggtcctcaaggaacagctac 1249  
DB 5128 gaacagacactgctgcagagagagagagagagagagagagagagagagagagag 5164

Search completed: May 20, 2000, 00:35:01  
Job time: 11884 sec

1	1288	99.2	1298	1	X40056	Colon cancer asso
2	191.2	14.7	6644	1	X33181	Base sequence of t
3	191.2	14.7	7372	1	X33182	Base sequence of t
4	191.2	14.7	7797	1	X33180	Cowpox virus bsr f
5	191.2	14.7	7996	1	X33184	Base sequence of t
6	148.2	11.4	1666	1	Q87587	DNA encoding Leuco
7	145	11.2	4000	1	T91902	Manose-1-phosphat
8	142.2	11.0	3399	1	T05668	Chicken leucocytoz
9	125.6	9.7	11901	1	X02998	Human IL-1ra BAC c
10	125.2	9.6	2056	1	V62142	HSV-2 scrlain SB5 C
11	124	9.6	11084	1	X22302	Human IL-1ra BAC c
12	118	9.1	6755	1	V21511	Human IL-1ra BAC c
13	117.6	9.1	8952	1	V62130	staphylococcal bac
14	116.2	9.0	44453	1	X23519	HSV-2 strain SB5 C
15	114.6	8.8	29392	1	V15422	Human kidney amino
16	112.6	8.7	700	1	V88807	Mouse poly Ig recec
17	110	8.5	11495	1	X53491	EST clone IAZ00, N
18	109.4	8.4	10620	1	X02996	Human adenosine A1
19	109.4	8.4	14650	1	X22303	Human IL-1ra BAC c
20	107.4	8.3	980	1	V06088	Human IL-1ra BAC c
21	103.8	8.0	53526	1	T94101	Viral infection ge
22	103.8	8.0	53577	1	T18551	Human PKD1 gene, H
23	103.8	8.0	53577	1	T94108	Human poly cystic k
24	99.8	7.7	12700	1	V62133	Human PKD1 locus b
25	98.8	7.6	5741	1	V62150	HSV-2 strain SB5 C
26	98.4	7.6	4550	1	V60472	HSV-2 strain SB5 C
27	98.2	7.6	9789	1	T41852	Sequence encoding
28	96.4	7.4	3653	1	X20280	CDNA encoding plas
29	95.4	7.3	19124	1	T72882	Borrelia burgdorfe
30	95	7.3	2784	1	T65102	Plasmodium var-7 g
31	93.6	7.2	240	1	T76782	Luciferase gene au
32	91.6	7.1	2447	1	V54587	Staphylococcus aur
33	91.6	7.1	35515	1	X20252	Human secretory pr
34	89.8	6.9	110000	1	V21209_06	Borrelia burgdorfe
						Continuation (7 of

[illegible]

## ALIGNMENTS

RESULT	1
X40056	
ID	X40056 standard; DNA: 1298 BP.
AC	X40056;
DT	02-JUL-1999 (first entry)
DE	Colon cancer associated gene.
KW	Cancer associated antigen; diagnosis; research; treatment; human; breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer; prostate cancer; ss.
OS	Homo sapiens.
PN	W03904265-A2.
PD	28-JAN-1999.
PF	15-JUN-1998; U14679.
PR	22-JUN-1998; US-102322.
PR	17-JUL-1997; US-896164.
PR	10-OCT-1997; US-061599.
PR	10-OCT-1997; US-061765.
PR	10-OCT-1997; US-948705.
PR	11-OCT-1997; GB-021697.
PI	(LUDMW-) LUDWIG INST CANCER RES.
PI	Chen Y, Gout I, Gure A, Ohare M, Obata Y, Old LJ, Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E, Tureci O;
PI	WTCL; 99-132448/11.
PT	New isolated cancer associated nucleic acids and polypeptides - isolated using sera from cancer patients, used to develop products for the diagnosis, monitoring or treatment of cancers
PS	Claim 67, Page 654-655; 787pp; English.
CC	The invention relates to a method for diagnosing a disorder characterised by expression of a human cancer associated antigen precursor coded for by a nucleic acid molecule (NAM). The method comprises: (a) contacting a biological sample isolated from a subject with an agent that specifically binds to the NAM, an expression product or a fragment of an expression product complexed with an HLA molecule; and (b) determining the interaction between the agent and the NAM or the expression product as a determination of the disorder. The products and methods can be used in the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens. The invention provides nucleic acid sequences and encoded polypeptides which are cancer associated antigen precursors expressed in human breast cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and lung cancer.
SQ	Sequence 1298 BP; 546 A; 189 C; 379 G; 174 T;
Query Match	99.2%; Score 1288; DB 1; Length 1298;
Best Local Similarity	100.0%; Pred. No. 2e-224;
Matches 1298:	Conservative 0; Mismatches 0; Indels 0
Y	1 ggcctgtaaatgactgcgaaccggcttcagagagcccttcgtcctttagacancanaga 60
D	1 ggctgctcaaatgactgcgaaccggcttcagagagcccttcgtcctttagacancanaga 60
O	61 agaacatgcgatcttgccaanaagactacctcttgatgtagaagtgaagaatgaaggggagaca 120
B	61 agaacttacggagattttgccaaaagaaactacctcttggatgtagaagtgaagaatgaaggggagaca 120





[illegible]

SOURCE	3
ID	X33182 standard; DNA; 7372 BP.
AC	X33182;
DT	25-JUN-1999 (first entry)
DE	Base sequence of the plasmid pRX-Bcl-xl-psr.
KW	Compx virus; bsr; viral vector; expression; apoptosis; resistance;
KW	crrna; bcl-2; bcl-xl, FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW	autoimmune disease; graft rejection reaction; inflammation;
KW	inflammatory disease; ss.
OS	Synthetic.
OS	Homo sapiens.
PN	MO9913073-A2.
PD	18-MAR-1999.
PF	07-SEP-1998; J04010.
PR	08-SEP-1997; JP-259235.
PA	(RPRG-) RPR GENCELL ASIA PACIFIC INC.
PI	Hamada H;
DR	WPI: 99-243728/20.
PT	New apoptosis-resistant virus-sensitive cell
PS	Example 2; Page 41-45; 51pp; English.
CC	The present invention describes an apoptosis-resistant virus-sensitive
CC	cell line into which an apoptosis resistance gene has been introduced.
CC	The recombinant viruses generated are capable of expressing apoptosis-
CC	associated genes. These can then be used in a variety of diseases for
CC	which the induction of apoptosis by gene transfer, or where the
CC	inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC	are useful as vectors for gene therapy which can be applied to cancer
CC	therapy for destroying cancer cells selectively, the treatment of
CC	autoimmune diseases and graft rejection reaction, and apoptosis induction
CC	therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC	encountered the problem where if an adenovirus vector capable of
CC	expressing an apoptosis-associated gene is introduced into animal cells,
CC	the cells producing the virus will be destroyed because the period of
CC	time required to induce cell death by apoptosis is shorter than that
CC	required to replicate and produce the virus, resulting in failure to
CC	obtain a recombinant virus having the integrated apoptosis-associated
CC	gene. In this invention an apoptosis-resistant 293 cell line (having an
CC	apoptosis resistant gene introduced) is established and overcomes the
CC	problem. The present sequence represents the base sequence of the
CC	plasmid pRX-Bcl-xl-psr, which contains the human Bcl-xl gene, and
CC	is used in an example from the present invention.
SQ	Sequence 7372 BP; 2353 A; 1749 C; 1649 G; 1621 T;

[illegible]

RESULT 4  
 X33180  
 ID X33180 standard; DNA; 7/97 BP.  
 AC X33180;  
 DT 25-JUN-1999 (first entry)  
 DE COMPOX virus bsr full length gene sequence.  
 KW COMPOX virus; bsr; viral vector; expression; apoptosis; resistance;  
 KW crmk; bcl-2; bcl-x1; FLIP; SURVIVIN; IAP; ILP; adenovirus; cancer;  
 KW autoimmune disease; graft rejection reaction; inflammation;  
 KW inflammatory disease; ss.  
 OS COMPOX virus.  
 PN W09913073-A2.  
 PD 18-MAR-1999.  
 PF 07-SEP-1998; J04010.  
 PR 08-SEP-1997; JP-259235.  
 PI (RPRG-) RPR GENCELL ASIA PACIFIC INC.  
 PI Hamada H;  
 DR WPI; 99-243728/20.  
 PT New apoptosis-resistant virus-sensitive cell  
 PS Example 1: Page 34-38; 51pp; English.  
 CC The present invention describes an apoptosis-resistant virus-sensitive  
 CC cell line into which an apoptosis resistance gene has been introduced.  
 CC The recombinant viruses generated are capable of expressing apoptosis-  
 CC associated genes. These can then be used in a variety of diseases for  
 CC which the induction of apoptosis is gene transfer, or where the  
 CC inhibition of harmful apoptosis is therapeutic. The recombinant viruses  
 CC are useful as vectors for gene therapy which can be applied to cancer  
 CC therapy for destroying cancer cells selectively, the treatment of  
 CC autoimmune diseases and graft rejection reaction, and apoptosis induction  
 CC therapy for inflammatory cells in inflammatory diseases. Prior arts have  
 CC encountered the problem where if an adenovirus vector capable of

CC expressing an apoptosis-associated gene is introduced into animal cells,  
CC the cells producing the virus will be destroyed because the period of  
CC time required to induce cell death by apoptosis is shorter than that  
CC required to replicate and produce the virus, resulting in failure to  
CC obtain a recombinant virus having the integrated apoptosis-associated  
CC gene. In this invention an apoptosis-resistant 293 cell line (having an  
CC apoptosis resistant gene introduced) is established and overcomes the  
CC problem. The present sequence represents the compox virus bsr gene which  
CC is used in an example from the present invention.  
SQ Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T;

Query Match 14.7%; Score 191.2; DB 1; Length 7797;  
Best Local Similarity 57.3%; Pred. No. 1.3e-26;  
Matches 340; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 706 tgcagctcactatgagccagagctcgagagagagaaatcnaaagttaaagttc 765  
DB 4883 TGCAGTTCGCAA 4942  
QY 766 acaagtcgtgaagaaagaaagccctaaagagtttgagcagctgcgga 825  
DB 4943 AA 5002  
QY 826 aggttaatccagctgcgcactagaagaaagaaagagagagagagaa 885  
DB 5003 AA 5062  
QY 886 agaagaaagaaagagagagagagagagagagagagagagagagag 945  
DB 5063 AA 5122  
QY 946 aggaagaaagaaagagagagagagagagagagagagagagagagag 1005  
DB 5123 AA 5182  
QY 1006 gaactaagaaagagagagagagagagagagagagagagagagagag 1065  
DB 5183 AA 5242  
QY 1066 aagaagaaagagagagagagagagagagagagagagagagagagag 1125  
DB 5243 AA 5302  
QY 1126 gaaagaaagaaagagagagagagagagagagagagagagagagag 1185  
DB 5303 AA 5362  
QY 1186 agaagaaag 1245  
DB 5363 AA 5422  
QY 1246 agcggaagaaagaaagagagagagagagagagagagagagagagag 1298  
DB 5423 AA 5475  
RESULT 5  
X33184  
ID X33184 standard; DNA; 7996 BP.  
AC X33184;  
DT 25-JUN-1999 (first entry)  
DE Base sequence of the plasmid pR-X-Bcl 2-1-hCD 25.  
KW Compo virus; bsr; viral vector; expression; apoptosis; resistance;  
KW ctmA; bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;  
KW autoimmune disease; graft rejection reaction; inflammation;  
KW inflammatory disease; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN W09913073-A2.  
PD 18-MAR-1999.  
PE 07-SEP-1998; J04010.  
PR 08-SEP-1997; JP-259235.

PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.  
PI Hamada H;  
DR WPI: 99-243728/20.  
PT New apoptosis-resistant virus-sensitive cell  
PS Example 3; Page 46-49; 51pp; English.  
CC The present invention describes an apoptosis-resistant virus-sensitive  
CC cell line into which an apoptosis resistance gene has been introduced.  
CC The recombinant viruses generated are capable of expressing apoptosis-  
CC associated genes. These can then be used in a variety of diseases for  
CC which the induction of apoptosis by gene transfer, or where the  
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses  
CC are useful as vectors for gene therapy which can be applied to cancer  
CC therapy for destroying cancer cells selectively, the treatment of  
CC autoimmune diseases and graft rejection reaction, and apoptosis induction  
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have  
CC encountered the problem where if an adenovirus vector capable of  
CC expressing an apoptosis-associated gene is introduced into animal cells,  
CC the cells producing the virus will be destroyed because the period of  
CC time required to induce cell death by apoptosis is shorter than that  
CC required to replicate and produce the virus, resulting in failure to  
CC obtain a recombinant virus having the integrated apoptosis-associated  
CC gene. In this invention an apoptosis-resistant 293 cell line (having an  
CC apoptosis resistant gene introduced) is established and overcomes the  
CC problem. The present sequence represents the base sequence of the  
CC plasmid pR-X-Bcl 2-1-hCD 25, which contains the human Bcl-2 gene, and  
CC is used in an example from the present invention.  
SQ Sequence 7996 BP; 2463 A; 2015 C; 1829 G; 1689 T;

Query Match 14.7%; Score 191.2; DB 1; Length 7996;  
Best Local Similarity 57.3%; Pred. No. 1.3e-26;  
Matches 340; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 706 tgcagctcactatgagccagagctcgagagagagaaatcnaaagttaaagttc 765  
DB 5082 TGCAGTTCGCAA 5141  
QY 766 acaagtcgtgaagaaagaaagccctaaagagtttgagcagctgcgga 825  
DB 5142 AA 5201  
QY 826 aggttaatccagctgcgcactagaagaaagaaagagagagagagagagag 885  
DB 5202 AA 5261  
QY 886 agaagaaagaaagagagagagagagagagagagagagagagagagag 945  
DB 5262 AA 5321  
QY 946 agaagaaagaaagagagagagagagagagagagagagagagagagag 1005  
DB 5322 AA 5381  
QY 1006 gaactaagaaagagagagagagagagagagagagagagagagagagag 1065  
DB 5382 AA 5441  
QY 1066 aagaagaaag 1125  
DB 5442 AA 5501  
QY 1126 gaaagaaagaaagagagagagagagagagagagagagagagagagag 1185  
DB 5502 AA 5561  
QY 1186 agaagaaagaaagagagagagagagagagagagagagagagagagag 1245  
DB 5562 AA 5621  
QY 1246 agcggaagaaagaaagagagagagagagagagagagagagagagagag 1298  
DB 5622 AA 5674

RESULT 6  
ID 087587 standard; DNA: 1686 BP.  
AC 087587;  
DT 19-DEC-1995 (first entry)  
DE DNA encoding Leucocytozoon protozoa structural protein epitope.  
KW leucocytozoon protozoa; structural protein; epitope; vaccine; fowl;  
KM leucocytozoonosis; treatment; ss.  
OS Leucocytozoon protozoa sp.  
PN J07089995-A.  
PD 04-APR-1995.  
PF 10-SEP-1993; 226078.  
PR 10-SEP-1993; JP-226078.  
PA (DOBU-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.  
PM (NISS-) NISSEIKEN KK.  
DR WPI; 95-167252/22.  
P-PSDB: R70491.  
PT Immune inducing polypeptide against leucocytozoon protozoa - useful  
in production of vaccines for treatment of leucocytozoonosis in  
fowl.  
PS Claim 1; Page 12-14; 20pp; Japanese.  
CC 087587-89 encode polypeptides having a whole or partial epitope of a  
structural protein of Leucocytozoon protozoa (see R70491-93). The  
CC polypeptides and DNA encoding them are useful in the production of  
CC vaccines for the treatment of leucocytozoonosis of fowl.  
SQ Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T;

Query Match 11.4%; Score 148.2; DB 1; Length 1686;  
Best Local Similarity 59.6%; Pred. No. 6.7e-19;  
Matches 317; Conservative 0; Mismatches 211; Indels 4; Gaps 4;

QY 769 aagtcgtgaagaagaaggaagcccaagaagccctaaagatttgacgctgcggaag 828  
DB 101 AAGAGAAATATGAGAAATGAAAAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAG 160  
QY 829 ttaatccgcgtccgcactagaagaacggaagaagaagaa-agaagggggggaag 887  
DB 161 AACCAAGAACAAATCGTAGAAGAACCAAGAACCAAGAACCAAGAACCAAGAAC 220  
QY 888 aagaagaacaaggaagaagaagaagaagaagaagaagaagaagaagaagaaga 947  
DB 221 ATGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 279  
QY 948 gaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1006  
DB 280 GAAGAAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 339  
QY 1007 aactaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1066  
DB 340 GATGAGAAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 399  
QY 1067 agaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1126  
DB 400 CAAATGAGATGACAAATGAAAGTGAACAAATGAAAGCAAGCAAGCAAGCAAG 459  
QY 1127 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1186  
DB 460 GAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 519  
QY 1187 gaaaga-aagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1245  
DB 520 GAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 579  
QY 1246 agcggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1297  
DB 580 AATGAAGATGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 631

RESULT 7  
T91902 standard; DNA: 4000 BP.  
ID T91902

AC T91902;  
DT 07-MAY-1998 (first entry)  
DE Mannose-1-phosphate transferase gene MNNA.  
KW Mannose-1-phosphate transferase; MNNA gene; enzyme; yeast; regulatory gene;  
human; high mannose type neutral saccharide chain; ss.  
OS Saccharomyces cerevisiae.  
FH Key location/Qualifiers  
FT CDS 418..3953  
FT /tag= a  
FT /product= MNNA

PN J0926792-A.  
PD 14-OCT-1997.  
PF 29-MAR-1996; 075667.  
PR 29-MAR-1996; JP-075667.  
PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.  
PM WPI; 97-553460/51.  
DR P-PSDB: W30763.  
PT Positive regulatory gene of mannose-1-phosphate transfer in yeast -  
useful for high mannose type neutral saccharide chain production  
PS Claim 2; Page 14-17; 23pp; Japanese.  
CC This sequence represents the gene of the invention, designated MNNA,  
CC encoding a protein which positively regulates mannose-1-phosphate  
CC transfer in yeast. The gene is useful for the preparation of human high  
CC mannose type neutral saccharide chain.  
SQ Sequence 4000 BP; 1453 A; 750 C; 770 G; 1027 T;

Query Match 11.2%; Score 145; DB 1; Length 4000;  
Best Local Similarity 65.9%; Pred. No. 2.7e-18;  
Matches 274; Conservative 0; Mismatches 128; Indels 14; Gaps 4;

QY 882 agaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 941  
DB 3527 AGAAACCGAGCGAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3586  
QY 942 gaagagaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 999  
DB 3587 GAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 3646  
QY 1000 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1059  
DB 3647 AAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3705  
QY 1060 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1119  
DB 3706 GGAGAAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3759  
QY 1120 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1179  
DB 3760 GATGAGAAATGAGATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAATG 3819  
QY 1180 agaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1239  
DB 3820 AGAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 3879  
QY 1240 gtagaagcggaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1295  
DB 3880 GAAGCAAGGAGG-----AGGAAAGAAAGAAAGCAAGCAAGCAAGCAAGCAAG 3930

RESULT 8  
T05868 standard; DNA: 3399 BP.  
ID T05868  
AC T05868;  
DT 14-AUG-1996 (first entry)  
DE Chicken leucocytozoon DNA encoding immunogenic protein for vaccines.  
KW Chicken leucocytozoon; immunogen; recombinant vaccine; protection;  
immunisation; vaccination; ss.  
OS Chicken leucocytozoon.  
FH Key location/Qualifiers  
FT cds 1..3399  
FT /tag= a  
FT misc\_feature 1150..3218

Query Match	11.0%	Score 142.2	DB 1	Length 3399
Best Local Similarity	57.6%	Pred. No. 8.5e-18		
Matches 328	Conservative	0	Mismatches 232	Indels 9
			Gaps	
739	agaagaaatcnaaagtttaaagatcaaaagtcgtgaagaagaagcgaag	798		
2595	AGAAAGAAAGAAAGAAAGAAAGAAAGTAAACATGAGAGAAAGAAAGTAAACATGAGAG	2654		
799	ccctaaagagtltagcagctgcggaagtgtaatccagctgcgcactagaagaag	858		
2655	AGAAAGAAAGAAAGTAAACATGAGAGAGAAAGAAAGTAAATATGAGAGAAAGAAAGAAAG	2714		
859	gaaagaggaagaagagagagagagaagaagaagaagaagaagaagaag	918		
2715	AGAAATGATGGGAG	2774		
919	ggaagaggaagaag-----aagaaggaagaaggaagaagaagaagaag	973		
2775	TGAGGAG	2834		
974	agaaggaagaagaagaagaagaagaagaagaagaagaagaagaagaaga	1032		
2835	GGAAAG	2894		
1033	agaaggaagaagaagaagaagaagaagaagaagaagaagaagaagaaga	1092		
2895	AGAAATGATGATGAG	2954		
1093	gaagacctaagaagaagaagaagaagaagaagaagaagaagaagaaga	1152		
2955	ACATGAG	3014		
1153	ggaagaagaagaataagaagaagaagaagaagaagaagaagaagaaga	1211		
3015	AGAAAGAAAGTAAACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	3074		
1212	gaaagaagaaga--aaaaaggaagaagaatagaagcgaagaagaagaagaat	1269		
3075	AACATATGAG	3134		
1270	aagaaggaagaagaagaagaagaagaagaagaagaagaagaagaaga	1298		
3135	GAGAGAGAGATATGAG	3163		

DE	Human IL-1ra BAC contiguous DNA sequence 43.
KW	Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;
KM	interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
KW	chronic myelogenous leukemia; psoriasis; inflammatory bowel disease;
KM	growth factors; treatment; IL-1 receptor complex; BAC; ss.
OS	Homo sapiens.
PN	WO906426-A1.
PD	11-FEB-1999.
PF	03-AUG-1998; U16102.
PR	02-JUL-1998; US-091650.
PR	04-AUG-1997; US-054646.
PA	(MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PI	Pan Y;
PT	WPI; 99-133692/13.
PT	New isolated nucleic acid encoding the new human cytokine Tango-77 -
PS	used to inhibit inflammation and to screen for specific modulators
PP	Example 5; Figure 3; 226BP; English.
CC	X02956-X03048 and X22301-X22304 are overlapping BAC genomic sequences
CC	containing alternatively spliced forms of human IL-1ra. Such fragments
CC	are used in the method of the invention which describes the isolation of
CC	a novel human TANGO-77 encoding nucleic acid and protein. Tango-77 is a
CC	member of the cytokine superfamily that is expected to inhibit
CC	inflammation by binding to the interleukin-1 receptor (IL-1R). It may
CC	also bind to a new receptor so could regulate other cellular processes
CC	associated with acute or chronic inflammation, e.g. asthma, chronic
CC	myelogenous leukemia, rheumatoid arthritis, psoriasis and inflammatory
CC	bowel disease. It may also induce or suppress interleukins, cytokines and
CC	growth factors. Modulators of this protein are used to treat or prevent
CC	conditions associated with abnormal levels of inflammation, or activity
CC	of IL-1 or its receptor complex.
SQ	Sequence 11901 BP; 3552 A; 2724 C; 2438 G; 3179 T;

Query Match	9.7%	Score 125.6;	DB 1;	Length 11901;
Best Local Similarity	61.1%;	Pred. No. 9.2e-15;		
Matches 218;	Conservative	0;	Mismatches 137;	Indels 2;
			Gaps	1
OY 853	aacgaagaaagaaggaagaaggaaggagaagaagaacaaggaagaagaag	912		
D6 6414	ACCAGAAAACAAGCAGAAGAGAGAGAGAAAGGAGAAAGAAAGAAAGAAAG	6473		
OY 913	aagaagggagaaggaagaagaagaagaagaagaagaagaagaagaagaag	972		
D6 6474	AAAGAAAGAAATGTGAAGGAAGGAAGGAAGAAAGAAAGAAAGAAAGCAAG	6533		
OY 973	gagaaggaagaagaagaagaagaagaagaagaagaagaagaagaagaata	1032		
D6 6534	AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGCGACGAA	6593		
OY 1033	agaaggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag--	1091		
D6 6594	ACGAGGAAGGAAGAAAGAAAGCAAGCAAGCAAGGAAGGAAGGAAGGAAGGA	6653		
OY 1091	aggaagaactnagaagaagaagaagaagaagaagaagaagaagaatnagaag	1150		
D6 6654	ACGAGGAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	6713		
OY 1151	aaggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga	1207		
D6 6714	AAAGGAAGAAAGAAAGAAAGGAAGGAAGGAAGGAAGAAAGAAAGCAAGAAAGAA	6770		

RESULT 10
V62142/c
ID V62142 standard; DNA: 2056 BP.
AC V62142:
DT 07-DEC-1998 (first entry)
DR HSV-2 strain SB5 Contlig ID 82 DNA sequence.
KM HSV-2 strain SB5: immunological response induction; therapy;
KW antiviral identification; viral protein inhibitor; ss.
OS Herpes simplex virus type 2.
FH key Location/Qualifiers
FT CDS 761..1150

QY	DB	Sequence	Best Local Similarity	Score	DB 1	Length	2056
QY	733	gaagagagaaagaaatcnaaagttaaaagtacaaagtcgtgaagaagaagaaagccca	9.6%	125.2	DB 1	2056	792
DB	1956	GGAGGAGAGAGGAGAAATAGAGAGTAAATAGAGAGAGGAGATGAGCTTAAATAGAGGGGA	51.4%	9.5e-15			1897
QY	793	agaaagccctaagaagttgagcgcgtcgcggaagtttaattccagctgcgcactagaag	0	274			852
DB	1896	AGGAGAGACGAGGAGAGAAAAAGAGGGGGGAGAAAGAGAGGGGGGAGCAACAGAGACA					1837
QY	853	aacgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga					910
DB	1836	GAG					1777
QY	911	aga					970
DB	1776	GAG					1717
QY	971	aggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga					1030
DB	1716	GGCAGCAG					1657
QY	1031	taaga					1090
DB	1656	AG					1597
QY	1091	aggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga					1150
DB	1596	AG					1537
QY	1151	aaga					1210
DB	1536	AAAG					1477
QY	1211	aggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga					1270
DB	1476	CCAG					1417
QY	1271	aga					1298
DB	1416	ACAG					1389

ID	Accession	Standard	DNA	11084 BP
AC	X22302			
DE	22-JUN-1999	(first entry)		
KM	Human IL-1ra BAC	contiguous DNA sequence 95		
KM	Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;			
KM	interleukin-1 receptor; IL-1R; psoriasis; asthma; rheumatoid arthritis;			
KM	chronic myelogenous leukemia; peristalsis; inflammatory bowel disease;			
KM	growth factors; treatment; IL-1 receptor complex; BAC; ss.			
OS	Homo sapiens.			
PN	MO906426-A1.			
PD	11-FEB-1999			
PF	03-AUG-1998;	U16102.		
PR	02-JUL-1998;	US-091650.		
PR	04-AUG-1997;	US-054646.		
PA	(MIL-) MILLENNIUM BIOTHERAPEUTICS INC.			
PI	Pan Y;			
DR	WPI. 99-153692/13.			
PT	New isolated nucleic acid encoding the new human cytokine Tango-77 -			
PT	used to inhibit inflammation and to screen for specific modulators			
PS	Example 5; Figure 4; 226pp; English.			
CC	X02956-X03048 and X22301-X22304 are overlapping BAC genomic sequences			
CC	containing alternatively spliced forms of human IL-1ra. Such fragments			
CC	are used in the method of the invention which describes the isolation of			
CC	a novel human TANGO-77 encoding nucleic acid and protein. Tango-77 is a			
CC	member of the cytokine superfamily that is expected to inhibit			
CC	inflammation by binding to the interleukin-1 receptor (IL-1R). It may			
CC	also bind to a new receptor so could regulate other cellular processes			
CC	associated with acute or chronic inflammation, e.g. asthma, chronic			
CC	myelogenous leukemia, rheumatoid arthritis, psoriasis and inflammatory			
CC	bowel disease. It may also induce or suppress interleukins, cytokines and			
CC	growth factors. Modulators of this protein are used to treat or prevent			
CC	conditions associated with abnormal levels of inflammation, or activity			
CC	of IL-1 or its receptor complex.			
SQ	Sequence 11084 BP; 3205 A; 2388 C; 2598 G; 2892 T;			
Query Match 9.6%; Score 124; DB 1; Length 11084;				
Best Local Similarity 61.3%; Pred. No. 1.8e-14;				
Matches 214; Conservative 0; Mismatches 133; Indels 2; Gaps 1				
QY	853 aacgaagaagaaggaagaaggaggaggaagaagaagaacaagaagaagaagaag	912		
DB	816 ACCAGAAAAAAGAAAGAGAGAGAGAGAGAAAGAGAAAGAAAGAAAGAAAG	757		
QY	913 aagaagggaagaagggaagaagaagaaggagggaagaagaagaagaagaagaag	972		
DB	756 AAGAAAGAAAGATGTGAAAGGAAGGAAGGAAGAAAGAAAGAAAGTGAAGAAAGAAAGAAAG	697		
QY	973 gagaaggaaagaagaagaagaagaagaagaagaactaagaagaaggagggaagaata	1032		
DB	696 AAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	637		
QY	1033 agaaaggaaagaagaagaagaagaagttnaaagaagaagaagaagaagaagaag--	1091		
DB	636 AGAAGAAAGGAAGAAAGAAAGAAAGCAAGCAAGCAAGCAAGAAAGAAAGCAAGCA	577		
QY	1091 aggaagaacttngaagaagaagaaggagggaagaagaagaagaagaataaagaacnagaag	1150		
DB	576 AGCAAGGAAGAAAGTAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	517		
QY	1151 aaggaagaagaagaataaagaagaaggagggaagaagaagaagaagaagaagaag	1199		
DB	516 AGAAAGAAAGAAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAG	468		
RESULT 12				
ID	V21511			
AC	V21511; standard; DNA; 6755 BP.			
DT	17-AUG-1998 (first entry)			
KM	Stephylococcal Bacteriocin BacRI operon.			
KM	BacRI; bacteriocin; antimicrobials; antibacterial; antibiotic;			











Query Match 3.6%; Score 81.4; DB 1; Length 7218;  
Best Local Similarity 4.7%; Pred. No. 1.1e-11;  
Matches 19; Conservative 246; Mismatches 142; Indels 0; Gaps 0;

QY 1032 gcaagcgagagcttcatacagaagcgctgctgagatgagtcacaagaatccca 1091  
DB 1438 GTACRRR 1379  
QY 1092 ggaagcagcagagatgagcgagcgaagagaagaatggccacagaagcagaaga 1151  
DB 1378 RR 1319  
QY 1152 aaatgagatataccggaagagatgagacagattgagagagaagaagttaaga 1211  
DB 1318 RR 1259  
QY 1212 gcaatgagagaagactgggctcaagaacagctactcttgccataaacctcagc 1271  
DB 1258 RR 1199  
QY 1272 tgaagtaacccagtaaccccttcgcagcgaagatgagtgagagtgagactagct 1331  
DB 1198 RR 1139  
QY 1332 cgaagccgagatgactgagatgagcagagagcagagagcagagattcccgaa 1391  
DB 1138 RR 1079  
QY 1392 atagagagaagcttgaccctactctatgtacccacagagaga 1438  
DB 1078 RR 1032

RESULT 2

US-08-728-323A-1  
Sequence 1, Application US/08728323A  
Patent No. 5948676  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA  
TITLE OF INVENTION: Encoding Same And Uses Thereof  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,323A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3489 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3489  
US-08-728-323A-1

Query Match 3.1%; Score 68.4; DB 4; Length 3489;  
Best Local Similarity 4.7.6%; Pred. No. 2e-08;  
Matches 201; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 979 ggcgttcatacagaccggagcgctgctgagagcgcgagagtgagctgacggcg 1038  
DB 2182 GAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2241  
QY 1039 cagagcttcatacagagcgctgctgagatgagtcacaagaatccctcagagagcag 1098  
DB 2242 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2301  
QY 1099 cagagatgagcgcgcaagaagaagaatgcccagaagcgagcagagagaatgag 1158  
DB 2302 GAGTTAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2361  
QY 1159 agataccggaagagatgagaaagatgtagagagagagagaagttaagaagcagtg 1218  
DB 2362 CAGCAGTTAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2421  
QY 1219 gaagaagactgggctcgaagagacagctactcttgcttaaaacccactgctgagta 1278  
DB 2422 GAGCAGAGATTAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2481  
QY 1279 caaccaagtaaccccttcgcagacgaagtatgatacagagtgagactgagagccc 1338  
DB 2482 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2541  
QY 1339 gcaagatgacctgtagagcagcgagcagagagcagagagcagagattccggaatctgag 1398  
DB 2542 GAGCAGAGAGTGGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2601  
QY 1399 ga 1400  
DB 2602 GA 2603

RESULT 3

US-08-770-379-20/C  
Sequence 20, Application US/08770379  
Patent No. 5849564  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED  
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

RESULT 5  
 US-09-130-114-1/c  
 Sequence 1, Application US/09130114  
 Patent No. 5976807  
 GENERAL INFORMATION:  
 APPLICANT: Horlick, Robert A.  
 APPLICANT: Damaq, Bassam B.  
 APPLICANT: Robbins, Alan K.  
 TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes  
 TITLE OF INVENTION: From Multiple Transfected Episomes

```

: FILE REFERENCE: 0867/ID9030US1
: CURRENT APPLICATION NUMBER: US/09/130,114
: CURRENT FILING DATE: 1998-08-06
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 5452
: TYPE: DNA
: ORGANISM: YEBNA
: US-09-130-114-1

```

Query Match	2.6%	Score 58	DB 4	Length 5452
Best Local Similarity	46.9%	Pred.No. 1.2e-05		
Matches 181	Conservative	0	Mismatches 205	Indels 0
			Gaps	0

QY	997	cgagagcgcgtctggcacaagagcgcgcgcaagcgtgaactctcaacggcacaagaattctcatgacg	1056
Db	1835	CAGGAGCAGCAGCAGGAGGCGCAGGAGCAGGAGGCGCAGGAGGCGCAGGAGGCGCAGGAGCAG	1776
QY	1057	aagcgcgtctggcgtatgtgaattccaacaagatctctccagagcaacagagaatgtgacgcgca	1116
Db	1775	GAGGCGCAGGAGCAGGAGGCGCAGGAGGCGCAGGAGGCGCAGGAGGCGCAGGAGGCGCAG	1716
QY	1117	agagcaaaagaaattctgccagaagcagcagagaaatgtagagataccggaagagatg	1176
Db	1715	GAGCAGGAGGCGCGCAGGAGCGAGGAGGCGCAGGAGCAGGAGGCGCAGGAGGCGCAGGAGCAG	1656
QY	1177	gaacagattgtagagggagagaagaagattaaagaacaaatggagaagaactcggggtcca	1236
Db	1655	GAGGCGCGCAGGAGGCGCGAGGAGCGAGGAGGCGCAGGAGGCGCAGGAGGCGCAGGAGGCGCAG	1586
QY	1237	aaggaacagctactctctgtccctaaacatcatctctgtcagtgatcacccagtaacctctcgc	1296
Db	1595	GAGGCGCGCAGGAGCGAGGAGCGAGGCGCGCAGGAGGCGCAGGAGGCGCGCAGGAGGCGCGAG	1536
QY	1297	aagccaaagtatgatcatcagggagttgaaacctgagcttcgagccgcagatgacctgatatga	1356
Db	1535	GAGCAGGAGGCGCGCAGGAGGCGCGAGGAGCGAGGAGGCGCGAGGAGGCGCGAGGAGGAGGAG	1476
QY	1357	ggcagcgagagcagggagagcgagga	1382
Db	1475	GCGCAGGAGCAGGAGGCGCGAGGAGCGCA	1450

RESULT 6  
US-07-884-811-15  
; Sequence 15, Application US/07884811  
; Patent No. 5316921

```

: NAME: Dreger, Ginger R.
: REGISTRATION NUMBER: 33,055
: REFERENCE/DOCKET NUMBER: 75.1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/225-3216
: TELEFAX: 415/952-9881
:
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 15:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 10596 bases
:
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-07-884-811-15

```

Query Match	2.6%	Score 58	DB 1	Length 10596
Best Local Similarity	46.9%	Pred. No. 1.7e+05		
Matches 181	Conservative	0	Mismatches 205	Indels 0
				Gaps 0

QY	997	cggaagcgcgtcgtgagagagcgcgacagcgttgaagctcagccggcagaagcttctatcag	1056
Db	2550	CAGGACGAGGAGGAGGGGGCAGAGACAGAGAGGGGCAGAGAGGGGCACGAGCAGC	2609
QY	1057	aagcgccttgcgahcggatcccaacaagatccctccaggaagcagcagagatgtagcgcgca	1116
Db	2610	GAGGGCGAGGAGCAGAGGAGGGGGCAGAGAGGGGCACAGAGGGGCACGAGAGGGCGAG	2665
QY	1117	aggaagaaagaatattgccagaagcagcagagaaatctagatataccggaagagatg	1178
Db	2670	GAGCAGAGAGGGGCGAGGAGCGAGGAGGGGGCAGAGACAGAGAGGGGCAGAGGGCGACGAGCAG	2739
QY	1177	gaacagattttagtaggagagagagaagatttaagaacaaatctggagaagaacctgggctca	1236
Db	2730	GAGGGGCGAGGAGGGGGCGAGGACGAGAGGGGCGAGAGAGGGGCAGAGAGAGGGCGAG	2788
QY	1237	aaggaacacagctactctctgccttaaacatcatcctgcgtgaagtatcacccagttacccctctgc	1296
Db	2790	GAGGGCGAGGAGCGAGGAGGAGGGGGCAGAGAGGGGCACAGACAGAGAGGGGCACGAGGGCGAG	2849
QY	1297	aagccaaagatattgatcagggagttgaaacctgagctcgagcccgacagatgatctgtgatg	1356
Db	2850	GAGCAGAGGAGGGCGAGGAGGGGGCAGAGAGCAGAGAGGGGCACAGAGGGGCAGAGCAGAGCAG	2909
QY	1357	ggccagcgagagcagggagagagcgga	1382
Db	2910	GGGCAGGAGCAGAGGGGGCGAGGAGCA	2935

RESULT 7  
 US-07-885-971-15  
 : Sequence 15 Application US/07885971  
 : Patent No. 5328837  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
 : TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS  
 : NUMBER OF SEQUENCES: 21  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Genentech, Inc.  
 : STREET: 460 Point San Bruno Blvd  
 : CITY: South San Francisco  
 : STATE: California  
 : COUNTRY: USA  
 : ZIP: 94080  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: 5.25 Inch, 360 Kb floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: patin (Genentech)  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/07/885,971  
 : FILING DATE: 19920518  
 : CLASSIFICATION: 530  
 :



```

; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,088B
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallegos, R. Thomas
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: 755D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-2614
; TELEFAX: 910/371-7168
; TELETYPE: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-194-088B-15

```

```

Query Match          2.6%; Score 58; DB 1; Length 10596;
Best Local Similarity 46.9%; Pred. No. 1.7e-05;
Matches 181; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

```

```

QY 997 cggagcggctgagcagagcgcgagcgtgagctgacgcgagcagagcttcctcagcag 1056
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2550 CAGGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2609
QY 1057 aagcgctgctgagtgagtgccaaagatccctcagagcagcagagagatgagcgagca 1116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2610 GAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2669
QY 1117 aggaagaagaatgccagaagcagcagaggaagaatgagagatcccggaagagagatg 1176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2670 GAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2729
QY 1177 gaacagattgtagaggaagaagaagtttaagaagaacatggaagaaagactgaggctca 1236
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2730 GAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2789
QY 1237 aaggaacagctactcttgcttaaacatcactgctgagtgatcaaccagtaacccttcgc 1296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2790 GAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2849
QY 1297 aagcacaatgatgataaggaagtgaaacctgagctcgagcccgagatgacctgagtgga 1356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2850 GAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2909
QY 1357 ggcacggaagcagcaggaagagcagga 1382
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2910 GGGCAGAGCAGAGGAGGAGGAGGAGCAGCA 2935

```

```

RESULT 10
US-08-194-087-15
; Sequence 15, Application US/08194087
; Patent No. 5879310

```

```

; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,087
; FILING DATE: 18-MAY-1992
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 910/371-7168
; TELETYPE: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-194-087-15

```

```

Query Match          2.6%; Score 58; DB 3; Length 10596;
Best Local Similarity 46.9%; Pred. No. 1.7e-05;
Matches 181; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

```

```

QY 997 cggagcggctgagcagagcgcgagcgtgagctgacgcgagcagagcttcctcagcag 1056
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2550 CAGGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2609
QY 1057 aagcgctgctgagtgagtgccaaagatccctcagagcagcagagagatgagcgagca 1116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2610 GAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2669
QY 1117 aggaagaagaatgccagaagcagcagaggaagaatgagagatcccggaagagagatg 1176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2670 GAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2729
QY 1177 gaacagattgtagaggaagaagaagtttaagaagaacatggaagaaagactgaggctca 1236
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2730 GAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2789
QY 1237 aaggaacagctactcttgcttaaacatcactgctgagtgatcaaccagtaacccttcgc 1296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2790 GAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2849
QY 1297 aagcacaatgatgataaggaagtgaaacctgagctcgagcccgagatgacctgagtgga 1356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2850 GAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2909
QY 1357 ggcacggaagcagcaggaagagcagga 1382
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2910 GGGCAGAGCAGAGGAGGAGGAGGAGCAGCA 2935

```

```

RESULT 11
PCT-US93-04648-15
; Sequence 15, Application PC/TUS9304648
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie

```



NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-104P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2793 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Nephila clavipes  
TISSUE TYPE: minor amputillate gland  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 183..2675  
OTHER INFORMATION: /product= "N. clavipes minor  
amputillate silk protein"  
OS-08-209-747-1

Query Match	2.2%;	Score 49.4;	DB 2;	Length 2793;
Best Local Similarity	43.7%;	Pred. No. 0.0016;		
Matches 107; Conservative	0;	Mismatches 138;	Indels 0;	Gaps 0;

QY	976	cggagagctgttcataagacaagcccgagagcgcgcctgcagagagcgccgacgctgaagctgcag	1033
Db	850	CTGAGACAGAGATCTGGAGAGCGGCTGCGCGTTACGGTGAAGGTCTGTGCTGGAGCTGGAG	909
QY	1036	cgcgcagagacttcctcatalgcagaaagcgcgctgycgatagagatccatccagag	1095
Db	910	CCGCTGCAGGTGCAGAGAGCGAGAGCTGGAAGCTACGCTGATCAAGGATACGGTGCAGGAG	969
QY	1096	cagcagagagatltgagagcgcgcaaaagagaataattgcgccagagcagcagagaaat	1155
Db	970	CAGGAGCTGGTGCTGCTGCAGCTGCANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	1023
QY	1156	gagagatacccggaagagagatgcagacagatltgataagagaaagagatlttaagaagcaa	1215
Db	1030	NNNNNNNNGGTCAGAGTGCAGTGTCTGATATGTGTGCACAGAGCGGATATGTGTCCGGAG	1089
QY	1216	tgagga	1220
Db	1090	CAGGA	1094

```

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046

```

```

1 APPLICATION NUMBER: US/08/458,298
2
3 FILING DATE: 02-JUN-1995
4
5 CLASSIFICATION: 530
6
7 PRIOR APPLICATION DATA:
8 APPLICATION NUMBER: US 08/209,747
9 FILING DATE: 14-MAR-1994
10
11 ATTORNEY/AGENT INFORMATION:
12 NAME: Murphy Jr., Gerald M.
13 REGISTRATION NUMBER: 28,977
14 REFERENCE/DOCKET NUMBER: 1447-104H
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE: 703-205-8000
17 TELEFAX: 703-205-8050
18
19 INFORMATION FOR SEQ ID NO: 1:
20
21 SEQUENCE CHARACTERISTICS:
22 LENGTH: 2793 base pairs
23 TYPE: nucleic acid
24 STRANDEDNESS: double
25
26 TOPOLOGY: linear
27
28 MOLECULE TYPE: cDNA
29
30 HYPOTHETICAL: NO
31
32 ORIGINAL SOURCE:
33 ORGANISM: Nephila clavipes
34 TISSUE TYPE: minor ampullate gland
35
36 FEATURE:
37 NAME/KEY: CDS
38 LOCATION: 183..2675
39 OTHER INFORMATION: /product= "N. c
40 OTHER INFORMATION: ampullate silk
41
42 US-08-458-298-1

```

```

; OTHER INFORMATION: /product="N. clavipes minor
; OTHER INFORMATION: ampullate silk protein"
US-08-458-298-1

```

Query Match	2.28;	Score 49.4;	DB 2;	Length 2793;
-------------	-------	-------------	-------	--------------

	Matches	107;	Conservative	0;	Mismatches	138;	Indels	0;	Gaps	0;
QY	976	cgggagcgtgtcatgacagaccgggagcggcgtgcagagagccgcgacagcgtgactgtcag								1035
Db	850	ctggagccagagatctggagagccgctgcgctgacggtagagagctctgctgtgagctggag								909
QY	1036	cggcagagagcttctcatgcagaagcggcgtgcgtatggaagtcacacaagaattccctcagag								1095
Db	910	ccgctgcagagctcagagagacgagcgtggaaagctagcgtgtgtcgaaggatracggtgcgcgag								969
QY	1096	cagcagagagatgtagacggtcgaagggagaaaaaagtctgccagaagcagcagagagaaat								1155
Db	970	caggagcgtggtctctctcagcctgcagcannnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn								1029
QY	1156	gagagataccggaagaagaatggacaagatgtcagaagggaagaagaagtttaagaagcaa								1215
Db	1030	nnnnnnnnngtctcagagtgacagtgctgtgatgatgtgcgacaaacggatgtgctgcccgag								1089
QY	1216	tgggga	1220							
Db	1090	cagga	1094							

RESULT 15  
US-08-056-200-93  
; Sequence 93, Application US/08056200  
; Patent No. 5616500  
; GENERAL INFORMATION:  
; APPLICANT: Steinert, Peter M.  
; APPLICANT: Lee, Seung-Chul  
; APPLICANT: Kim, In-Gyu  
; APPLICANT: Chung, Soo-Il  
; APPLICANT: Park, Sang-Chul  
; TITLE OF INVENTION: Trichopyalin and Transglutaminase-3 and  
; TITLE OF INVENTION: Methods of Using Same  
; NUMBER OF SEQUENCES: 117  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor



```

1 CITY: Newport Beach
2 STATE: CA
3 COUNTRY: U.S.A.
4 ZIP: 92660
5
6 COMPUTER READABLE FORM:
7 MEDIUM TYPE: Floppy disk
8 COMPUTER: IBM PC compatible
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10 SOFTWARE: Patentin Release #1.0, Version #1.25
11
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/056,200
14 FILING DATE: 30-APR-1993
15 CLASSIFICATION: 435
16
17 ATTORNEY/AGENT INFORMATION:
18 NAME: Fedrick, Michael F.
19 REGISTRATION NUMBER: 36,799
20 REFERENCE/DOCKET NUMBER: NIH054,001A
21
22 TELECOMMUNICATION INFORMATION:
23 TELEPHONE: (714) 760-0404
24 TELEFAX: (714) 760-9502
25
26 INFORMATION FOR SEQ ID NO: 93:
27
28 SEQUENCE CHARACTERISTICS:
29 LENGTH: 9551 base pairs
30 TYPE: nucleic acid
31 STRANDEDNESS: single
32 TOPOLOGY: linear
33
34 MOLECULE TYPE: cDNA
35 HYPOTHETICAL: NO
36 ANTI-SENSE: NO
37
38 FEATURE:
39 NAME/KEY: CDS
40 LOCATION: 1507..1644
41
42 FEATURE:
43 NAME/KEY: Intron
44 LOCATION: 1645..2511
45
46 FEATURE:
47 NAME/KEY: CDS
48 LOCATION: 2512..8070
49
50 US-08-056-200-93

```

Query Match	2.2%	Score 48.2:	DB 1:	Length 9551:
Best Local Similarity	48.4%:	Pred. No. 0.0057:		
Matches 134:	Conservative	0:	Mismatches 143:	Indels 0:
				Gaps 0:
QY 973	gaccggagagctgttatcaacagaccggagcgctctggcaagcgcgagcgctgagctg	1032		
Db 4888	ggcccccacgcacgcgcgagaggagaaagagcttgcagtctcttgagagaaagagacgcctc	4947		
QY 1033	cagcgcgagagagcttctcatgacagagcgcgctggatgagatccacaagaatctccag	1092		
Db 4948	cagcgcgcgagagcctgcccacacagctccagagagagagagagagcgcctccagagagatcag	5007		
QY 1093	gagcgcagagagatgtagcgcgagcaaaagagaaaagaatthgccagaaagcagcaagaa	1152		
Db 5008	gagagagagcgacagagagacgagcgccgcaccaaaatgtagggtgggaactagaaagaa	5067		
QY 1153	aatgagagataccgagaaagagatgtaaacagatctgtagagagagaaagaagtttaagaa	1212		
Db 5068	aggaaagacacgcgcacacacgcctgttaccgcgaacccctcaagaagcagcttgaggaag	5127		
QY 1213	caatgggaagaagacttggtggtccaaagaacgctac	1249		
Db 5128	gaacagcagctectctcagagagagagagagagctac	5164		

Search completed: May 20, 2000, 00:33:02  
Job time: 11765 sec

**THIS PAGE BLANK (USPTO)**

1	2236	100.0	2236	1	X39644	Renal cancer assoc
2	2236	100.0	2236	1	X40057	Colon cancer assoc
3	2173	97.2	2289	1	X40061	Colon cancer assoc
4	2078	92.9	2162	1	X40058	Colon cancer assoc
5	2053	91.8	2409	1	X40062	Colon cancer assoc
6	1306	58.4	1306	1	X40060	Colon cancer assoc
c	68.4	3.1	32207	1	V73801	KSHV LTR unique C
7	68.4	3.1	137507	1	V119941	KSHV LTR unique C
8	58	2.6	799	1	V55831	Nucleotide sequence
9	58	2.6	9600	1	V21683	Vector plasmid pCMV
10	58	2.6	10596	1	051731	Plasmid pCISEBON f
11	58	2.6	10596	1	051731	Plasmid pCISEBON f
12	58	2.6	10596	1	051731	Plasmid pCISEBON f
13	58	2.6	10596	1	051731	Plasmid pCISEBON f
14	57	2.5	795	1	V55830	FluA insert stable
15	52	2.3	543	1	023092	Antigen tc-7a gene
16	49.4	2.2	2744	1	098470	MSP1-containing F
17	48.8	2.2	29392	1	058422	Mouse poly Ig rece
18	47.6	2.1	4000	1	T91902	Mannose-1-phosphat
19	46.8	2.1	2338	1	014183	N. clavipes draglin
20	46.8	2.1	2338	1	V23249	N. clavipes draglin
21	46.4	2.1	2000	1	N71065	Gene encoding Plas
22	46	2.1	243	1	037291	Clone 2-9 used in
23	46	2.1	243	1	037291	SSP 7.7.7.7.7.8
24	46	2.1	243	1	035004	Synthetic storage
25	46	2.1	243	1	035004	Synthetic storage
26	46	2.1	1390	1	V84574	Human secreted pro
c	45.6	2.0	12001	1	076213	HSV L/ST region, f
28	45.4	2.0	234	1	084832	Spino cerebellar at
29	45.2	2.0	1505	1	055750	Genomic clone G1H4
30	45.2	2.0	12301	1	V20445	Human c-erbB oncog
c	44.6	2.0	117213	1	V62146	HSV-2 strain SB5 C
32	44	2.0	1686	1	087587	DNA encoding leucoc
33	44	2.0	1995	1	014184	N. clavipes draglin
34	44	2.0	1995	1	V23250	N. clavipes draglin

C 35	4.4	2.0	3512	1	X37085	EMF associated proc
C 36	43.6	1.9	154	1	Q84835	Spinoecerebellar at
C 37	43.6	1.9	203	1	V17226	SCA2 gene CAG repe
C 38	43.6	1.9	203	1	V30271	Glutamine rich reg
C 39	43.6	1.9	913	1	T07199	Cotton fibre-speci
C 40	43.6	1.9	913	1	T13034	Cotton fibre-speci
C 41	43.6	1.9	913	1	T30253	Cotton fibre cell
C 42	43.6	1.9	913	1	T62610	Cotton fibre speci
C 43	43.6	1.9	913	1	T70041	Cotton fibre speci
C 44	43.6	1.9	1984	1	T13030	Cotton fibre-speci
C 45	43.6	1.9	1985	1	T30250	Cotton fibre clone

## ALIGNMENTS

RESULT	1
ID	X39644
AC	X39644 standard; DNA; 2236 BP.
DT	02-JUL-1999 (first entry)
DE	Renal cancer associated gene.
KW	Cancer associated antigen; diagnosis; research; treatment; human; breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer; prostate cancer; SS.
OS	Homo sapiens.
PN	W0904265-A2.
PD	28-JAN-1999.
PF	15-JUL-1998; U14679.
PR	22-JUN-1998; US-102322.
PR	17-JUL-1997; US-896164.
PR	10-OCT-1997; US-061599.
PR	10-OCT-1997; US-061765.
PR	10-OCT-1997; US-948705.
PR	11-OCT-1997; GB-021697.
PA	(LUDWIG-) INST CANCER RES.
PI	Chen Y, Gout I, Gure A, OHare Y, Old LJ, Pfeunderschuh M, Sahin U, Scanlan MJ, Stockert E, Tureci O.
PI	WPJ; 99-132448/11.
PT	New isolated cancer associated nucleic acids and polypeptides -
PT	isolated using sera from cancer patients, used to develop products
PT	for the diagnosis, monitoring or treatment of cancers
PS	Claim 67; Page 435-436; 787pp; English.
CC	The invention relates to a method for diagnosing a disorder characterised
CC	by expression of a human cancer associated antigen precursor coded for by
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC	biological sample isolated from a subject with an agent that specifically
CC	binds to the NAM, an expression product or a fragment of an expression
CC	product complexed with an HLA molecule; and (b) determining the
CC	interaction between the agent and the NAM or the expression product as a
CC	determination of the disorder. The products and methods can be used in
CC	the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens.
CC	The invention provides nucleic acid sequences and encoded polypeptides
CC	which are cancer associated antigen precursors expressed in human breast
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC	lung cancer.
CC	Sequence 2236 BP; 548 A; 588 C; 653 G; 447 T;

Query Match	100.0%;	Score 2236;	DB 1;	Length 2236;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2236; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

[illegible]

QY 121 ttcgcgacaagtgatgattctctgatltgaataatgatcgaggaaggactatctatgat 180  
|||||  
Db 121 TTCGGGCATTAAGTGATGATTTCTTGATTTGAAAATGATGCAGAGAGACATATCTCTATGAT 180  
QY 181 gtgtcgtgaatgtaccacagacacatlgagcgtgtgcgtgtcgttggagaaacttgaaagt 240  
|||||  
Db 181 GTGTCTGCGATGTACCCACAGACCAATGAGACGTGGCCGTCTCTGTGGAGACCTGGAAGCTG 240  
QY 241 gtcatcaatgaaccccgagcgtctgtcctctgtttgatgtcattcgtgcgcgtgataccaaagt 300  
|||||  
Db 241 GTCATCAATGAACCCAGCCGTCTGCTCTGTGTTGATGCCATTTGGCCGCTGATCCCACTG 300  
QY 301 aagcaccaagtggaatataatcatcagctgcaccccgcgctccaggaagcttgaaagagtgtc 360  
|||||  
Db 301 AAGCACCAAGTGGAAATATGATGATGACGTGACCCCGCGCTCCAGAACTGTAAGGAGGTG 360  
QY 361 cgtctgagaccgtctgcaccccggaagcctcgcgcgtgagtgtcgtgtgtgtgtgtgtgtgtt 420  
|||||  
Db 361 CGTCTGAGACCGTCTGCACCCCGGAAGCCCTCGGCTGAGTGTGCTGTGGCTGGAGTTT 420  
QY 421 ggctgtgggtcttcaatctccacactcatcaaaagcggtataggcagacagcgtcgaggtc 480  
|||||  
Db 421 GGCTGTGGGCTCTTCACTCTCCACTCATCAAAAGCGGTCAGGACACACGCTCGGAGCTC 480  
QY 481 caggttagggagcagatcgtccgagatcaatlgatlatcatctcctcctgtacccaatgag 540  
|||||  
Db 481 CAGGTAGGGAGAGATCGTCCGGATCAATGATATTCCATCTCTCTGTACCATGAG 540  
QY 541 gaggtatcaacctcatctgcgaaccaaagaactgtgtlccatcaaatlgagaacatcgcgc 600  
|||||  
Db 541 GAGGTATCAACCTCATTTGGAACCAAGAAACGTGTCCATCAAAATGAGACACATCGG 600  
QY 601 cgtatcccgtaaaaagcctctctgtatgaacccctcactgtgagatgttgaatcagttt 660  
|||||  
Db 601 CTGATCCCGTGAAGGCTCTCTGTATGAGCCCTCATCTTGGAGATGTGATTCAGTTT 660  
QY 661 gtgtcgtgaatcttggggcgttgcgagcagccttgcgtcctcccttggaaatcgggaaacaag 720  
|||||  
Db 661 GTGTGGAATCTGGGGCGTGGGAGCAGCCTGGGCTCCCTGGAATGCGGAAACAG 720  
QY 721 gagaagaagttcttcaatcagcctgtgtatagctcccgagagccttgcgtgcagcatlccagc 780  
|||||  
Db 721 GAGAAAGAGTCTATCATGACCCGTGAGGCTCCGAGGCTTGGCTGCACGATTTCCAG 780  
QY 781 gggcccatcagaagcctgcatcttcatcagacatgtigaacacttgcctcgtctgtct 840  
|||||  
Db 781 GGGCCCATCAAGAGCCTGGCATCTTATCAAGCATGTGAACCTTGGCTCCGTCTGCT 840  
QY 841 gaggttggatctggaatagggagacagatgttgcgaatcgaatggcgtlccatctctaac 900  
|||||  
Db 841 GAGGTGGATGGAGATGAGGGGACAGATTTGTGAATGATGGCTCGACTTCTCTAAC 900  
QY 901 ctggatcaaaaagagagctgtataatgtgtctgtaaaaaatagccgcagccttgaccatctcatt 960  
|||||  
Db 901 CTGGATCAACAAAGAGGCTGTAAATGTGCTGAATAATAGCCGACGCTGACCATCTCATT 960  
QY 961 gtaagtcgaagcttggcggagagctgttcatcagacagacggagcggttggcaagagcgcg 1020  
|||||  
Db 961 GTAAGTCGACGCTGGCCGGAGCTGTTTCATGACAGACCGGAGGCGGCTGGCAGAGGCGCG 1020  
QY 1021 cagcgttgaagcttgcagcggcaggaagcttctcatgtcagaagcggttggcgtatgagttcaac 1080  
|||||  
Db 1021 CAGCGTGAGCTGCGAGCGGACGAGGCTTTCATGATGCAAGAGCGGCTGGCGATGAGTCCAAC 1080  
QY 1081 aagatctctccagagcagcagagatgtgagcgycgcaagaggaataaatgtgccagaag 1140  
|||||  
Db 1081 AAGATCTCTCAGAGCAGCAGAGGATGAGCGGCAAGGAAATAATTTGCCCGAAG 1140  
QY 1141 gcagcagagaaatagagatatccggaagagatlgaaacagatgttagaggaagagag 1200  
|||||  
Db 1141 GCAGCAGAGAAATAGATAGATACCGGAAGAGATGGAACAGATTGTAGAGGAGGAAGAG 1200  
QY 1201 aagtttaagaagaatlggaagaagaacttgggtctcaagaagacagctactcttgcctaaa 1260

|||||  
Db 1201 AAGTTTAAGAGCAATGGGAAGAGACTGGGCTCAAAAGAAACAGTACTCTTGCTTAA 1260  
QY 1261 accataactgcttgaagttacacccagtaacctcttcgaagccaaagtatgatccaggaagt 1320  
|||||  
Db 1261 ACCATACGTCTGAGGTACACCCAGTACCCCTTTCGGAAGCAAAAGATATGATCGGAGATG 1320  
QY 1321 gaacctgaagctcgaagcccgagatgaacttgaatgtgagtgagcgagagagcagaggaagcag 1380  
|||||  
Db 1321 GAACCTGAGCTCGAGCCCGAGATGACGTGATGAGAGGACGAGGAGGAGGAGGAGGAGCAG 1380  
QY 1381 gatttccggaataatgaagaaagcttgaacctactatgtltaaccccaagacatc 1440  
|||||  
Db 1381 GATTTCGGAATAATGAGGAAGGCTTTGACCCCTACTATGTTCAACCCAGAGCAGATC 1440  
QY 1441 atgggaaagatgtccggtcctcagcatcaaaaagagagatccttgaagcgtgcctgt 1500  
|||||  
Db 1441 ATGGGAAGGATGTCCGGCTCTTACCATCAACAAAGAGGATCTTAGACCTTGACCTG 1500  
QY 1501 gaagcgtgttgaactccccatltggaaagtgtcgttctgttgtatgagcgggga 1560  
|||||  
Db 1501 GAAGCGGCTGTGACCTCCCATTTGGGAAGTGTGTCTCTGTATGAGCGGGGA 1560  
QY 1561 gctgtcgaagcgagcatgttggcatlgtgaaaggggacagatcatggaatcaacgycagag 1620  
|||||  
Db 1561 GCTGTGAGCGGCATGTGGCATTTGAAAAGGAGAGATCATGGAATCAACGCGCAAG 1620  
QY 1621 atgtgtacagactacacccctgtgtgagcgtgcgcgccttgaagaagccttgaatcag 1680  
|||||  
Db 1621 ATGTGTACAGACTACACCCCTGAGGCTGAGCGCTGCCCTGAGAAAGGCTTGGAATCAG 1680  
QY 1681 ggcgggagacttgaatcgaacttgtgtgtcgtctgcgctcccccgaagagatgaacgaatgag 1740  
|||||  
Db 1681 GCGGGGAGACTGAGTGCACCTGTTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740  
QY 1741 ctgaactcttctgaagtcgaatcaaaaggggaaccaaattcaacggttgaagaacagttag 1800  
|||||  
Db 1741 CTGACCTTCTTGTGTAAGTCCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800  
QY 1801 ctccggccccaactctgtgaacacaaagcctcgaaacagccttgaagaagggcacaatgaa 1860  
|||||  
Db 1801 CTCGGGCCCACTCGTGAACACAAAGCCTCGGACAGCCTTGAGAGAGGACCATGATGAC 1860  
QY 1861 caccacagatgcatccttggagaccttgaatctatcaacccagaaatcacaactcccttgc 1920  
|||||  
Db 1861 CACACAGATGGCATCTTGGGACCTGGAATCTATFCAACCAAGAACTCAAACTCCCTTTC 1920  
QY 1921 gccctgaaccagggccagataaagaaacagctcgggccacttlttgaagggcaatgttga 1980  
|||||  
Db 1921 GCCCTGAACCCAGGCGCAAGATTAAGGAACAGCTCGGGCACCTTTTGTGAMGGCCAAATGTGA 1980  
QY 1981 ggaagaagagcagcagcagcttgggaagaagatcctaagaatccagactcaatcttctt 2040  
|||||  
Db 1981 GGAAGAAGGACACGACGCCCTTGGGAGAGATCTCAAGGATCAAGATCTCAATTCCTTT 2040  
QY 2041 cctctggccagtgaaatttggltctctccagccttggggagactccttcttgaaccttaa 2100  
|||||  
Db 2041 CCTCTGGCCAGTGAATTTGGTCTCTCCAGCTTGGGGAGCTCTCTTGAACCTTAA 2100  
QY 2101 taagaccacatgtgagttctctctctccatccctctctctgcctctgccttcaattgc 2160  
|||||  
Db 2101 TAAAGACCCACAGTGAATCTCTCTCTCAATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2160  
QY 2161 tggcagatgtgtacatccaaacttactctcgtgagctatataataaataacagatattat 2220  
|||||  
Db 2161 TGGCAGGATTTGTACTTCAAACTTACTCTGAGCTATTAATAATAAACAATTTATTT 2220  
QY 2221 ttcagcttaaaaaa 2236  
|||||  
Db 2221 TTCACAGCTTAAAAAA 2236

RESULT 2

X40057  
ID X40057 standard; DNA; 2236 BP.  
AC X40057;  
DE 02-JUL-1999 (first entry)  
KW Colon cancer associated gene.  
KW Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
KW prostate cancer; ss.  
OS Homo sapiens.  
PN MO9904265-A2.  
PD 28-JAN-1999.  
PR 15-JUL-1998; 014679.  
PR 22-JUN-1998; US-102322.  
PR 17-JUL-1997; US-896164.  
PR 10-OCT-1997; US-061599.  
PR 10-OCT-1997; US-061765.  
PR 11-OCT-1997; US-948705.  
PR 11-OCT-1997; GB-021697.  
PA (LUDW-) LUDWIG INST CANCER RES.  
PI Chen Y, Gout I, Gure A, O'Hare M, O'bata Y, Old LJ,  
PI Preundschuh M, Sahin U, Scanlan MJ, Stockert E,  
PI Tureci O.  
DR MPI; 99-132448/11.  
PT New isolated cancer associated nucleic acids and polypeptides -  
PT isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
PS Claim 67, Page 655-656; 787pp; English.  
CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
SQ Sequence 2236 BP; 548 A; 588 C; 653 G; 447 T;

Query Match 100.0%; Score 2236; DB 1; Length 2236;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctggcccgctgcggtgcggtctcttcagctcctgcgagccgggacccgaagaac 60  
DB 1 CTTGGCCCGGTCCGCGTCCGCGGCTCTTCCAGCTCCTGGCAGCCGGGACCCGAGAGAC 60

QY 61 gggctgtcaacagcagcagcttgacctgcccagccatlgaccgaaagtggccggaa 120  
DB 61 GGGTGTCAACAGCAGCAGCTGGACCTGGCCGACCATGACCGAAAGTGGCCCGAGAA 120

QY 121 ttcgcgataaagtgatgttcttgatgtgaaatgatgacagaagacatctctatgat 180  
DB 121 TTTCCGATATAGGTGATTTTCTGATGTGAAATGATGACAGAGAGACTATCTTATGAT 180

QY 181 gtctctcgaatgtfacccagacacatgagctgagctgctgtgtggaagccgtgaagctg 240  
DB 181 GTCTCTCGAATGTGACCAACACCATGAGAGCTGCTGTGCGGAGACCTGAACCTG 240

QY 241 gtcatcaatgaaccacagctctgcctctgtttgatgcatctgcgcgctgacatccactg 300  
DB 241 GTCATCAATGAACCAACCGCTCTGCTCTGTTCATGCACTTCGCGCGCTGATCCCACTG 300

QY 301 aagcaccaagtggaataatgacagctgaccccccgcgcgtccaggaagctgaaagagtg 360  
DB 301 AAGCACCAAGTGGAAATGATGACGTGACCCCGCGCGCTCCAGAGACTGAAGAGAGTG 360

QY 361 cgtctgacgcgtctgcaccccggaagcctgcgctgagtggtggtgacctggaattc 420

DB 361 CGTCTGACCGGTGACACCCGAAAGGCGCTGAGTGTGCGTGGCTGGAATTT 420  
QY 421 ggtctgtggtctctcatctccacatcaaaagcggtcagagacagcgtcgtc 480  
DB 421 GGCTGTGGGCTCTTCATCTCCACCTCATCAAAAGCGGTGAGCAGACAGCTGGGCTC 480

QY 481 caggttagggagagatctccgagataatgatatccatctctctcttaccagag 540  
DB 481 CAGGTAGGGAGAGATCTCCGATCAATGGAATTTCCATCTCTCTGACCAATGAG 540

QY 541 gaagctcaaacctcatctgaaccagaagaaactgtgtcattcaaaagtgaacacatcgc 600  
DB 541 GAGCTCATCAACCTCATCTGACACCAAAACTGTGTCCATCAAAAGTGAACATATGCGC 600

QY 601 ctgattcccgatgaagaagctcctctgtagagccctcacttgcagatagtgtacagtt 660  
DB 601 CTGATCCCGGTGAAGAACTCTCTGATGAGCCCTCTACTTGGCAGTATGTGATCATGTT 660

QY 661 gtgtcggaaatctggggcgtgcagagcagcctgggctccctggaaatcgggaaacaag 720  
DB 661 GTGTGGAATCTGGGGCGTGCAGGACGCTGGGCTCCCTGGAAATCGGAAACAG 720

QY 721 gaagaagaagctctcatcaagcctgtgtagctcccgagccttgctgacatctcagc 780  
DB 721 GAGAAAGAAGCTTTCATCAGCTGTAGGCTCCCGAGGCTTGGCTGCACATTTCCAGC 780

QY 781 ggcctccatcaagaagcctgcatcttatacagcattgaaacctgtcctctgtct 840  
DB 781 GGCCCATTCAGAAAGCTGGCATCTTTATCAGCATGTGAACCTGCTCCCTGTCTGCT 840

QY 841 gaggttgatgtgagatagggagacaaatgttgaatcaatgctgcagatctctaac 900  
DB 841 GAGGTGGATGTGAGATAGGGAGACCAATGTCCAAATCAATGCGTGCATCTCTTAAC 900

QY 901 ctgattacaagaagagctgttaaatgtgtcgtgaaatatgcgcagcctgacatctcatc 960  
DB 901 CTGATTACAAAGAGGCTGTAAATGTGTCTGAAAAATATGCCGACCTGACCATTTCCATT 960

QY 961 gtagctgcaagctgcccggagctgttcatgacagaccggagcggctgcaagagcgcg 1020  
DB 961 GTAGCTGACAGCTGGCGGGAGCTGTTCATGACAGCCGGGAGCGGCTGGCAGAGGCGCG 1020

QY 1021 cagcgtgagcttgcagcgagcgagcttctcatcagaagcgctgctgagatgagatccaac 1080  
DB 1021 CAGCGTAGCTGCGCGGAGGACTTCTCATGAGAAAGCGGCTGGATGAGATCTCAAC 1080

QY 1081 aagatctccagagagcagcagagatgagcggcaagaggaagaaagaatgtgcccagaag 1140  
DB 1081 AAGATCTCCAGAGAGCAGCAGAGATGAGACGGCAAAAGAGAAATATGCCCAAGAG 1140

QY 1141 gcagcagaagaaatgaaatgaaatccggaagagatgaaacagatgtgtagagaaagag 1200  
DB 1141 GCAGCAGAGAAATGAGAGATACCGAAGAGATGGAACAGATTGTAGAGAGAGAGAG 1200

QY 1201 aagtttaagaagcaatgagaaagactggggtcctcaaggaacagctactcttgctaaa 1260  
DB 1201 AAGTTTAAAGAGCAATGGAAGAAAGACTGGGGCTCAAAAGAACAGCTACTCTTGCTAAA 1260

QY 1261 aacctcaactgtgaggtgacacccagtaaccccttcgcaagccaagtatgttcagagagtg 1320  
DB 1261 ACGATCACTGCTGAGGTACCCAGTACCCCTTCGCAAGCCAAAGTATGTAGACGGAGTG 1320

QY 1321 gaacctgagctgcagcccgagatgacctgtgtagagcagcggagagcagggagagacag 1380  
DB 1321 GAACCTAGCTGCAGCCCGAGATGACCTGGATGAGACGACGAGAGAGAGAGACAG 1380

QY 1381 gatttccggaataatgagaaagccttgaccctactatctatgttaccacagacagatc 1440  
DB 1381 GATTTCGGAATAATGAGAAAGCCTTGTACCCCTACTCTATGTTCACCCAGAGACAGATC 1440

QY 1441 atggggaagagatctcgcgctccctacgataaagaagaagagatctccttagagccgtg 1500

Db 1441 AAGGGAAGAGATGTCCGGCTCTACGACATCAGAAGAGAGATCTTGAAGCTGAGCCCTG 1500  
Qy 1501 gaagcggtgtgagaccgcccatctgggaaggtgtgctgtctgtgtatgagcgagga 1560  
Db 1501 GAAAGGGGTGTGACTCCCATTTGGGAAGGTGTCTTCTCTGTGTATGAGCGGGGA 1560  
Qy 1561 gctgtcgaagcagatgtgtgcatctgtgaaaggagagatcatggaatcaacgagcaag 1620  
Db 1561 GCTGTGAGCGGCGATGTGGCATTTGAAAGGGGAGAGATCATGCAATCAACGGCAAG 1620  
Qy 1621 attgtgaacagatcaacacctgtgtgagtgagtgagtgagtgagtgagtgagtgag 1680  
Db 1621 ATGTGTACAGACTACACCTCTGTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1680  
Qy 1681 ggcgggagatgagatgagatgagatgagatgagatgagatgagatgagatgagatgag 1740  
Db 1681 GCGGGGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1740  
Qy 1741 ctgacctctgtctgaagtcacaaaggagagagagagagagagagagagagagagagag 1800  
Db 1741 CTGACCTTCTGCTGAAGTCCAAAGGGGAAACCAATTCACGCTTGAAGAACAGTAGAG 1800  
Qy 1801 ctccggccccacctgtgtgacacaaagcctcgagacagccttgagagagagcagagaca 1860  
Db 1801 CTCCGGCCCCACTCTGTGAACACAAAGCCTCGGACCAAGCTTGAAGAGGCCACATGACA 1860  
Qy 1861 caaccagatgagatcctctgtgagacctgaatctatcaaccagagatctcaaacctcttg 1920  
Db 1861 CACACAGATGGATCTTGTGGACCTGGAATCTATCACCCAGAAATCTCAAACTCCCTTGTG 1920  
Qy 1921 ggcctgaacagagggccagataaagaaacagctcgagccactcttctgaaggccaatgtgag 1980  
Db 1921 GGCCTGAACAGAGGGCCAGATAAGAAAGACAGCTCGGGCCACTTTTGAAGGCCAATGTGAG 1980  
Qy 1981 ggaagagggagcgccagcgcttgggagagagatctcaagatccaaactctatctctt 2040  
Db 1981 GGAAGAGGGAGCACCGCGTTGGAGAAAGATCTCAAGATCCACACTCTCTCTTCTTCT 2040  
Qy 2041 cctctggccagatgaattgtgtctctcccaagcttgggagacccctctctctgaacctta 2100  
Db 2041 CCTCTGGCCAGTGAATTTGTGTCTCTCCAGACTTTTGGGGACTCTCTCTTGAACCTTAA 2100  
Qy 2101 taagaccacatgagatctctctctctctcatccctctctctctctctctctctctct 2160  
Db 2101 TAAAGACCCACATGAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2160  
Qy 2161 tgcagagatgtgacatcccaacttaactctgagctcatataataataacagattat 2220  
Db 2161 TGCAGAGATGTCTACTCAAAACCTTACTGTAGCTATTAAATTAACAGATTATTTAT 2220  
Qy 2221 ttccagcttaaaaaa 2236  
Db 2221 TTCCAGCTTAAAAAAA 2236

## RESULT 3

4X0061  
ID X40061 standard; DNA; 2289 BP.  
AC X40061:  
DT 02-JUL-1999 (first entry)  
DE Colon cancer associated gene.  
KW Cancer associated antigen; diagnosis; research; treatment; human;  
KW breast cancer; colon cancer; gastric cancer; lung cancer;  
KW prostate cancer; ss.  
OS Homo sapiens.  
PN W0904265-A2.  
PD 18-JAN-1999.  
PF 15-JUL-1998; U14679.  
PR 22-JUN-1998; US-102322.  
PR 17-JUL-1997; US-896164.  
PR 10-OCT-1997; US-061599.  
PR 10-OCT-1997; US-061765.  
PR 10-OCT-1997; US-948705.

PR 11-OCT-1997; GB-021697.  
PA (LUDW-) LUDWIG INST CANCER RES.  
PI Lunden Y, Gout I, Gure A, Ohare M, Obara Y, Old LJ,  
PI Pfeundschn M, Sahin U, Scanlan MJ, Stockert E,  
PI Tureci O.  
DR WPI; 99-132448/11.  
PT New isolated cancer associated nucleic acids and polypeptides -  
PT Isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
PS Claim 67, Page 663-664; 787pp; English.  
CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
SQ Sequence 2289 BP; 567 A; 596 C; 668 G; 458 T;

Query Match 97.2%; Score 2173; DB 1; Length 2289;

Best Local Similarity 97.7%; Pred. No. 0;

Matches 2236; Conservative 0; Mismatches 0; Indels 53; Gaps 1;

Qy 1 cctggccggtcggctcggctccttcagatcctctgagcagcgagcccgaaagaaac 60  
Db 1 CCGGGCCGGTCCGGTCCGGCTCTTCCAGCTCTGAGCGCGGGCACCCCAAGAAC 60  
Qy 61 gggctgtcaacagcagcagctgagcctggccagacatgagccgaagatgagccagaa 120  
Db 61 GGGCTGTCAACAGCAGCAGCTGAGCCTGGCCAGCATGAGCAGAAAGTGGCCCAAGAA 120  
Qy 121 ttccgcataaagtgatatttctgattgaaatgagcagagagagatctctatgt 180  
Db 121 TTCCGCATTAAGTGAATTTCTGATTGAAAAGATGACAGAGATCTATCTATGAT 180  
Qy 181 gtcgtcgaatgtaccacacagacatgagcgtgtgcctgtctgtgagacatcgaagctg 240  
Db 181 GTCTCGAATGTACACACAGACATGAGCTGCGCTGCTGCGGAGACCTGGAAGCTG 240  
Qy 241 gtcatcaatgaacccagcgtctcgtctgtgtgcatctgcgcgtgacccatg 300  
Db 241 GTCATCAATGAACCCAGCGCTGCTGCTGTTGATGCCATTCGGCGCTGATCCCATG 300  
Qy 301 aagcaccaagtgatataatgacagctgagcccgctccagagagatgagagagtg 360  
Db 301 AAGCACCAAGTGAATATATATACGTGACCCCGCGCTCCAGAGCTGAAGAGGTG 360  
Qy 361 cgtctgagcgtctgaccccgaaagcctcgagcctgagtgagtgagtgagtgagtgag 420  
Db 361 CGTCTGACCGTCTGTGACCCCGAAGCGCTGAGTGTGCTGTGTGCTGTGCTGTGCTGT 420  
Qy 421 ggcgtgtgagccttcaatcccaactcaatcaaaagcgtgacagcagcgtcgagctc 480  
Db 421 GCGTGTGGGCTCTTCAATCTCCCACTCATCAAAAGGGGTGAGCAGACAGCGTGGGCTC 480  
Qy 481 caggttagggagagatgctgagatcatatcatctccctcctgtaacatgag 540  
Db 481 CAGGTAGGGAGAGATGCTCGGATCAATGATATTCATCTCTCTGATCCCATGAG 540  
Qy 541 gaggtcatcaacctcatctgaaccaagaagaacatggtcatcaaaagtgaacacatcgag 600  
Db 541 GAGGTATCAACCTCTTGTGAACCAAGAAATGTTGATCAATCAAAAGTGAACATCGCG 600  
Qy 601 ctgattcccgtaaaagcctctcgtatgagccctcaacttgcagatgtgagatcagtt 660  
Db 601 CTTGATCCCGTAAAGCCTCTCTGATGAGCCCTCACTTGCAGATGTGAGATCAGTT 660

Db 601 CTGATCCCGTGAAGAGCTCTCTGATGAGCCCTCATTGCGAGTATGTGATCAGTTT 660  
QY 661 ggtgtggaatcttgaggcgctgagagcagcttggtctcccttggaatctgggaataaag 720  
Db 661 GTGTGGAATCTGGGGCGCGAGGAGCTGGCTCCCTCGAAATCGGAAACAG 720  
QY 721 gaagaagaagctcttcacacccgtgaagctcccgaggcttggtctgcaatcttcagc 780  
Db 721 GAAAGAAAGGTCTTTCATCAGACCTGTGAGGCTCCCGAGGCTTGGCTGACATTTCCAGC 780  
QY 781 ggcacacacagaagcctggcatcttalcacagcatgtgaacctggtccctgtctgt 840  
Db 781 GGGCCCATCCAGAGACCTGGCATCTTTATCAGGCATGTGAACCTGGCTCCGTGTGCT 840  
QY 841 gagggtggatcttgagatagggagaccagatctgtcgaagtcaatggcgtcgaactctaac 900  
Db 841 GAGGTGGGATGGAGATAGGGGACAGATGTGAAATGCAATGGCGCGCTCTCTAC 900  
QY 901 ctggaacacaaggagctgttaaatgtctgaataaatagccagcagccttgaccatctccatt 960  
Db 901 CTGGATCACAAGAGAGCTGTAAATGTGCTGAAAAATAGCCGACGCTGACCATCTCCATT 960  
QY 961 gtaagctcagcttgccggagctgtctcaltgaacagaccggagcgctgagagagcg 1020  
Db 961 GTAGCTGACAGCTGGCCGGAGCTGTTTCATACAGACCGGAGGCGCTGGCAGAGGCGCG 1020  
QY 1021 cagcgttgagctcagcgagcgagagctctcaltgcaagaacgctgctgagatgagctaac 1080  
Db 1021 CACCGTGAGCTGCAGCGGCGAGGAGCTTCATCAGAAAGGCGCTGGGAGTGGATCCAC 1080  
QY 1081 aagatctctcagaagcagcagagatctgagcgagcaagaggaagaatctgcagag 1140  
Db 1081 AAGATCTCTCAGAGAGAGAGCGAGGAGATGAGCGGCAAAAGGAAATAGGCTCCAGAG 1140  
QY 1141 gcagcagagaaatgaagataccggaagagatggaacagatctgagaagagagag 1200  
Db 1141 GCAGCAGAGGAAATGAGATACCGGAAGAGATGGAACAGATTGTAAGAGAGAGAG 1200  
QY 1201 aagtttaagaagaatgaggaagaagacttggtctcaagaagaacagctactcttgctaaa 1260  
Db 1201 AAGTTTAAGAAGCAATGGGAAAGACTGGGCTCAAAAGGAAACGATCACTTGGCCCTAAA 1260  
QY 1261 accatcactgctgaggtacaccagtaacctctgcagcagaagtatgtaacaggagtg 1320  
Db 1261 ACCATTACTGCTAGAGTACCCAGTACCCCTTCGCAAGCAAAATGATGATCGAGAGTG 1320  
QY 1321 gaacctgagctcagaccgagatgaacctgagtgagagcagcagagagagcag 1380  
Db 1321 GAACCTGAGCTCGAGCCGAGATGACCTGATGAGAGCGACGAGGAGAGAGAG 1380  
QY 1381 -----gattcc 1387  
Db 1381 CCACAGAGAGATGTTGAAGAGATGTGTTTATCAAGACAGATTCAGACAGATTTC 1440  
QY 1388 ggaataataagaaggcttgaccctactctatgttcaaccagaagagatctaggga 1447  
Db 1441 GGAATAATGAGAGGCTTTGACCCCTACTCTATGTTTACCCAGACCAATATATGGGGA 1500  
QY 1448 aggatgtccggctctcagcatcaagaagaaggagatccttagaaccttgccctggaagcg 1507  
Db 1501 AGGATGTCCGGCTCTACGATCAAGAGAGGAGATCCTTAGACCTGGCCCTGGAAGCG 1560  
QY 1508 gttgtgactcccccattgggaagtggtcttctcgtgtatgtagcgaggagctgtctg 1567  
Db 1561 GTGTGACTCCCATTTGGGAAAGTGTCTTTCTGCTGTGTGAGCGGGAGCTGCTG 1620  
QY 1568 agggagatgttgcatgtgaaggagagagatcatgtgaatcaacaggaagaattgtga 1627  
Db 1621 AGCGGATGTGCTATGTGAAGGAGACAGATCATGGGAATCAACGGCAAGATTGTGA 1680  
QY 1628 cagactacacccctgctgagagctgagcgtgcccctgcagaagcctggaatcagagcgagg 1687  
Db 1681 CAGACTACACCCCTTGCTGAGAGCTGAGCGCTCCCTGGAGAGGCTTGGAATCAGGCGGGG 1740

QY 1688 actgatacgaacctgtgtgtccgtctgcccccccaagagatagagatgagctgacct 1747  
Db 1741 ACTGATGACCTTGTGTGCTGCGTCCGCCCAAGGAGTGTGAGAGTGTGACCT 1800  
QY 1748 tctgtctgaagctccaaaggaggaaccaatctcagcggttaggaagaagtgagctccggc 1807  
Db 1801 TCTTGTGGAAGTCCAAAGGAGGAAACCAAAATTCACGGGTTAGGAACAGTGAAGCTCCGGC 1860  
QY 1808 cccaactcgtgaacacaaagcctcggagccagccttgagagagggccacatgacacacca 1867  
Db 1861 CCCACCTGTAACACAAAGCTCGAGACAGCCTTGAGAGAGGCGACATGACACACACA 1920  
QY 1868 gatgacatccttgaggactgaatctatcaccacgaagatctcaactcctttggcctga 1927  
Db 1921 GATGGCATCTCTTGGGACCTTAATCTATCACCGAGAAATCTCAAACTCCCTTGGCCCTGA 1980  
QY 1928 accagggccagataaagaagcctcggagccactttttgaaagccaaatgtgagagaaag 1987  
Db 1981 ACCAGGCCAGATTAAGAACAGCTCGGCGCATTTTGTGAAGGCCAATGTGAGGAAAG 2040  
QY 1988 gagcagccagccglttggaagaatctcaaggatccagactctcaltctcctctg 2047  
Db 2041 GAGCAGCCAGCCGTTTGGAGAAAGATCTCAAGGATCCAGACTGTCAATTCCTTCTCTGG 2100  
QY 2048 cccagtgaaattgtctctccagccttgaggagctcctccttgaaacctataaagacc 2107  
Db 2101 CCCAGTAAATTTGGTCTCTCCAGCTTGGGGAGACTCTTCTCTTGAACCTTAATTAAC 2160  
QY 2108 ccaactgagctctctctccatccctcctcctcctcctcctcctcctcctcctcctcct 2167  
Db 2161 CCACTGAGAGTCTCTCTCTCTCCATCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2220  
QY 2168 attgtcactcacaaccttactctgagctcatataaataaacaagattatcttcagc 2227  
Db 2221 ATTGTCACTCAAAACCTTACTGTGAGCTCATTTAATAAATAACAGATTATTTCAGC 2280  
QY 2228 ttaaaaaa 2236  
Db 2281 TTAATAAAAA 2289

RESULT 4  
X40058  
ID X40058 standard; DNA; 2162 BP.  
AC X40058;  
DT 02-JUL-1999 (first entry)  
DE Colon cancer associated gene.  
KW Cancer associated antigen; diagnosis; research; treatment; human;  
breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
prostate cancer; ss.  
OS Homo sapiens.  
PN M09904265-A2.  
PD 28-JAN-1999.  
PE 15-JUL-1998; U14679.  
PR 22-JUN-1998; US-102322.  
PR 17-JUL-1997; US-896164.  
PR 10-OCT-1997; US-061599.  
PR 10-OCT-1997; US-061765.  
PR 10-OCT-1997; US-948705.  
PR 11-OCT-1997; GB-021697.  
PA (LUDW-) LUDWIG INST CANCER RES.  
PI Chen Y, Gout I, Gure A, Ohare M, Ohta Y, Old LJ,  
PI Pfrendrich M, Sahin U, Scanlan MD, Stockert E,  
PI Tureci O;  
DR MPI: 99-132448/11.  
PT New isolated cancer associated nucleic acids and polypeptides -  
PT isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
PS Claim 67; Page 658; 787pp; English.  
CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a





OY	1921	gccctcgaaacagggccagataaagaacagctcgggccaatttlttgaaaggccaatgtga	1980
Db	1847	GCCCTGAACCAAGGGCGCAGATAAAGAACAAGCTGCGGCCACTTTTGTGAAGCCAATGTGA	1906
OY	1981	gaaagaggagacagcacgcgtttgggaagaagatctcaagaatccagactcatcttc	2040
Db	1907	GGAAAGGGAGCAGCCAGCCGTTTGGAAGANAGATTCAMAGATCCAGACTCTATTCCITTT	1966
OY	2041	cctctgcccacgtgaatttgtctctcccaagctttggggagactccttccttaaacctaa	2100
Db	1967	CCTCTGGCCCACTGAATTTGGTCTCTCCACGATTTGGGGGAGACTCCGTCTTAACOCCTNA	2026
OY	2101	taagaccaccaatcgtagagtctctctctccatccctctcctctgacctgtcctaayc	2160
Db	2027	TAAAGACCCACACGGAGGTCTCTCTCTCATCCCTCTCCCTGCCTTGCTCTTAATTC	2086
OY	2161	tgcagagattgcatctccaactctacctctgagctcatcataataaataaagattatt	2220
Db	2087	TGCCAGGATTTGCACTCCAAACCCTTACTCTGAGCTCATTAATAAATAAACAGATTATTT	2146
OY	2221	ttccagcttaaaaaa	2236
Db	2147	TTCCAGCTTAAAAAAA	2162
RESULT	5		
X40062			
ID	X40062 standard; DNA; 2409 BP.		
AD	X40062:		
DT	02-JUL-1999 (first entry)		
DE	Colon cancer associated gene.		
KW	Cancer associated antigen; diagnosis; research; treatment; human;		
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;		
KW	prostate cancer; ss.		
OS	Homo sapiens.		
PN	WO9904265-A2.		
ED	28-JAN-1999.		
PF	15-JUN-1998; U14679.		
PR	22-JUN-1998; US-102322.		
PR	17-JUL-1997; US-896164.		
PR	10-OCT-1997; US-061599.		
PR	10-OCT-1997; US-061765.		
PR	10-OCT-1997; US-948705.		
PR	11-OCT-1997; GB-021697.		
PI	(LUDMW-) LUDMIG INST CANCER RES.		
PI	Chen Y, Gout I, Gure A, Ohare W, Obata Y, Old LJ,		
PI	Pfeundschuh M, Sahin U, Scanlan MJ, Stockert E,		
PI	Tureci O;		
PI	WPI: 99-132448/11.		
PT	New isolated cancer associated nucleic acids and polypeptides -		
PT	isolated using sera from cancer patients, used to develop products		
PT	for the diagnosis, monitoring or treatment of cancers		
PS	Claim 67, Page 665-666; 787bp; English.		
CC	The invention relates to a method for diagnosing a disorder characterised		
CC	by expression of a human cancer associated antigen precursor coded for by		
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a		
CC	biological sample isolated from a subject with an agent that specifically		
CC	binds to the NAM, an expression product or a fragment of an expression		
CC	product complexed with an HLA molecule; and (b) determining the		
CC	interaction between the agent and the NAM or the expression product as a		
CC	determination of the disorder. The products and methods can be used in		
CC	the diagnosis, monitoring, research, or treatment of conditions		
CC	characterised by the expression of various cancer associated antigens.		
CC	The invention provides nucleic acid sequences and encoded polypeptides		
CC	which are cancer associated antigen precursors expressed in human breast		
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and		
CC	lung cancer.		
SQ	Sequence 2409 BP; 595 A; 638 C; 692 G; 484 T;		
Query Match	91.8%; Score 2053; DB 1; Length 2409;		
Best Local Similarity	92.8%; Pred. No. 0;		
Matches 2236; Conservative	0; Mismatches 0; Indels 173; Gaps 1;		

QY	1	ccgagcccgatcgcgatcgcgatctttccagctctctgycagccgggcaaccocgaagAAC	60
Db	1	cctggcccgcgatcgcgatcgcgatctttccagctctctgycagccgggcaaccocgaagAAC	60
QY	61	gggtctgagcaacgcgcgaagcttgagacccctggcccgacatgagccgaagaatggcccgagaa	120
Db	61	gggtctgagcaacgcgcgaagcttgagacccctggcccgacatgagccgaagaatggcccgagaa	120
QY	121	ttccggcataaggttgcatacttcttgatgtgaatatgatgatcagcagaagagactatctcatgat	180
Db	121	ttccggcataaggttgcatacttcttgatgtgaatatgatgatcagcagaagagactatctcatgat	180
QY	181	gtgctgcaatgatacacaacagaccatgagacgtgctgctgctggtggagacctgaaacgtg	240
Db	181	gtgctgcaatgatacacaacagaccatgagacgtgctgctgctggtggagacctgaaacgtg	240
QY	241	gtcatcaatgagaccacacgcttgccctctgtttgatgcatctgcgcgtatactccactg	300
Db	241	gtcatcaatgagaccacacgcttgccctctgtttgatgcatctgcgcgtatactccactg	300
QY	301	aagcaaccaggtgagataatgatacagctgaccccccgcgctccaggaagctgaagaagtgtg	360
Db	301	aagcaaccaggtgagataatgatacagctgaccccccgcgctccaggaagctgaagaaggtg	360
QY	361	cgctcgtgacgctgtgcaaccccggaagcctgcgctgagtgtgctgtgtgctgagatt	420
Db	361	cgctcgtgacgctgtgcaaccccggaagcctgcgctgagtgtgctgtgtgctgagatt	420
QY	421	ggctgtgagcctcttcatctctccacctcacaaggggtcagcgagacagcgtctgagctc	480
Db	421	ggctgtgagcctcttcatctctccacctcacaaggggtcagcgagacagcgtctgagctc	480
QY	481	caggtgaggggagcagatcgtcccgatcaatgatatctcatctccctccgtataccatgag	540
Db	481	caggtgaggggagcagatcgtcccgatcaatgatatctcatctccctccgtataccatgag	540
QY	541	gaagttcatcaaccctcatctcgacaaccaagaagaactgtgtccatcaaaagtgaacacatcgcc	600
Db	541	gaagttcatcaaccctcatctcgacaaccaagaagaactgtgtccatcaaaagtgaacacatcgcc	600
QY	601	ctgatcccgatgaaagcctctcctbatagacccctcaacttggcagatagtgatcagttt	660
Db	601	ctgatcccgatgaaagcctctcctbatagacccctcaacttggcagatagtgatcagttt	660
QY	661	gtgtcgtgaattgggggcgtgagcgagcgagcgtggtgtccctcgggaatccgggaacaaag	720
Db	661	gtgtcgtgaattgggggcgtgagcgagcgagcgtggtgtccctcgggaatccgggaacaaag	720
QY	721	gagaagaaggtctcatcacaacccgtgtagtcgtcccgagcgtctgtgctgacatctccagc	780
Db	721	gagaagaaggtctcttcatcacaacccgtgtagtcgtcccgagcgtctgtgctgacatctccagc	780
QY	781	ggccccatcagaagccttgacatcttatacagcatgtgaaactgtgctccctgtctgct	840
Db	781	ggccccatcagaagccttgacatcttatacagcatgtgaaactgtgctccctgtctgct	840
QY	841	gaagttgagattggagataaggggacccagattgttcaagtaagtaagtggttgacttctctaac	900
Db	841	gaagttgagattggagataaggggacccagattgttcaagtaagtaagtggttgacttctctaac	900
QY	901	ctgagatcaagaagggcgtgtgtaaatgtgtcgtgaatatagccgagcctgtgcatctccatt	960
Db	901	ctgagatcaagaagggcgtgtgtaaatgtgtcgtgaatatagccgagcctgtgcatctccatt	960
QY	961	gtagctgcagcttgcgcggagcgtgttcatgatcagacacgggagccggtgtgacagagcgcg	1020
Db	961	gtagctgcagcttgcgcggagcgtgttcatgatcagacacgggagccggtgtgacagagcgcg	1020
QY	1021	cagcgttgagctgcgcggcgagggccttctcatgtcagaaagcggtgtgcatgtgattccaac	1080
Db	1021	cagcgttgagctgcgcggcgagggccttctctatgtcagaaagcggtgtgcatgtgattccaac	1080

QY 1081 aagatccccaagagcagcagagatgagcgcgcaaaagagaaagaaattgccagaag 1140  
 DB 1081 AAGATCCTCCAGAGAGCAGCAGAGATGAGCGCAAAAGAGAAATTTGCCAGAG 1140  
 QY 1141 ggcagagagagaaatagagataccggaaagagatggaacagatttgaagagagaaag 1200  
 DB 1141 GCGAGAGAGAGAAATAGAGATACCGGAGAGAGATGAGAACATTTGAGGAGAGAGAG 1200  
 QY 1201 aagtttaagaagaatgagagagagagactggggtccaaaggaacagactcttgctaa 1260  
 DB 1201 AAGTTTAAAGAGAGATGAGAGAGAGACTGGGGCTCAAAAGAGAACAGCTACTTGGCTAAA 1260  
 QY 1261 accatcactgctgaggtacacccagtaacccttcgcgaagcaagtaagtaagtcaggagtg 1320  
 DB 1261 ACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCAAGTATGATCAGGGAGTG 1320  
 QY 1321 gaacctgagctgagcccgagagatgactgtgagtgagagcagagagagagagagagcag 1380  
 DB 1321 GAACCTGAGCTGAGCCCGAGATGACTGTGATGAGAGCAGGAGAGAGAGAGAGCAG 1380  
 QY 1381 ----- 1381  
 DB 1381 ACATTTTGGCCCAAGCCACAGCCTCCACAGGCCCTGGCTGTCCACATCTCCAAACT 1440  
 QY 1381 ----- 1381  
 DB 1441 GTCATGTCACACAGAGCCCAATTCATCTACAGGCAGCTGTGAATCTGAAGTTCTG 1500  
 QY 1381 -----gattcc 1387  
 DB 1501 CCACAGAGATGTTGAAGAGATGGTTTATCAAGACAGCATTCAGACAAAGATTTC 1560  
 QY 1388 ggaatatgagagagccttgagccctactctatgttcaaccagagagagatcatgagga 1447  
 DB 1561 GGAATAATGAGAGAGCCTTTGACCCCTACTCTATGTTTCAACCCAGAGCAGATATGAGGA 1520  
 QY 1448 aggaatgccgctcctcagcatcaagaagagagagatccttagaactgagccctgagagagc 1507  
 DB 1621 AGGATGTCGGGCTCTACGATCAAGAGAGAGGAGATCTTAGACCTGGCCCTGGAAGGCG 1680  
 QY 1508 ggtgagatcccccatttgagagagagtggtcttctgtgtgtatgagcggagagcgtcgt 1567  
 DB 1681 GGTGAGATCCCCCATTTGGAGAGGTGCTGCTTCTGCTGTGATGAGCGGAGCGGCTGTG 1740  
 QY 1568 agcgagatggtgagcatgtgaaagagagagatcatggaatcaacgagaaatgtgta 1627  
 DB 1741 AGCGGATGGTGGCATTTGTAAAGGGAGAGATCATGGAATCAACGCGCAATTTGTGA 1800  
 QY 1628 cagactacaacctgctgagagctgagcgtgcctgcaagaagcctggaatcaagggcgag 1687  
 DB 1801 CAGACTACACCCCTGGCTGAGGCTGAGCGTGCCTGCAAGAGCCCTGGAATCAGGGCGGGG 1860  
 QY 1688 actgagtgagcctgt 1747  
 DB 1861 ACTGGATCAGCCTTGT 1920  
 QY 1748 tcttgtagaagtcacaaagggagaaacaaatcaacgcttgaagaacagtgagctcgagc 1807  
 DB 1921 TCTTGCTGAAGTCCAAAGGGAGAACCAATTCACGCGTTAGAGAAACAGTACGACTCCGGC 1980  
 QY 1808 cccactcgtgaaacacaaagcctcgagcagccttgagagagggcacaatgacacacaca 1867  
 DB 1981 CCCACTCGTAACACAAAGCCTCGGACACGCTTGAGAGAGGCCACATGACACACACCA 2040  
 QY 1868 gatgagatccttggagcagcgaatctatcacccaggaatctcaactcccttggccctga 1927  
 DB 2041 GATGGATCCTTGGAGCCTGGAATCTATTCACCCAGGAATCTCAAACTCCCTTGGCCCTGA 2100  
 QY 1928 accaagggcagagatgaagaacagctcgagcagccttcttgaagggcaatgagagagag 1987  
 DB 2101 ACCAGGGCCAGATGAGGAACAGCTCGGGCCACTTTTGTAGGCCCATGTGTGGAGGAAGG 2160  
 QY 1988 gagcagcagcagcgttggagagagatctcaagatccagactctatcttcttctctg 2047

DB 2161 GAGCAGCAGCCGTTGGAGAGATCTCAAGATCCAGACTCATCTTCTCTCTGG 2220  
 QY 2048 cccagtgaaattgtctctccagcgttgggagcctctctcttgaaccttaagacc 2107  
 DB 2221 CCGAGTGAATTTGGTCTCTCCAGCTTTGGGGAGCTCTTCCCTTGAACCCTAATAGACC 2280  
 QY 2108 ccactgagctctctctctccatccctctcctctcctctcctctctctctctctctcag 2167  
 DB 2281 CCACTGGAGCTCTCTCTCTCCATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCAGG 2340  
 QY 2168 attgtcactcacaaccttactctgagctcaataaataaacaagattatcttcagc 2227  
 DB 2341 ATTGTCACTCAAACTTACTCTGAGCTCATTAATAAATAACAGATTATTTCCAGC 2400  
 QY 2228 ttaaaaaa 2236  
 DB 2401 TTAATAAAAA 2409

RESULT 6  
 X40060  
 ID X40060 standard; DNA; 1306 BP.  
 AC X40060;  
 DT 02-JUL-1999 (first entry)  
 DE Colon cancer associated gene.  
 KW Cancer associated antigen; diagnosis; research; treatment; human; breast cancer; colon cancer; gastric cancer; lung cancer;  
 KW prostate cancer; ss.  
 OS Homo sapiens.  
 PN W0904265-A2.  
 PD 28-JAN-1999.  
 PF 15-JUL-1998; U14679.  
 PR 22-JUN-1998; US-102332.  
 PR 17-JUL-1997; US-886164.  
 PR 10-OCT-1997; US-061599.  
 PR 10-OCT-1997; US-061765.  
 PR 10-OCT-1997; US-948705.  
 PR 11-OCT-1997; GB-021697.  
 PA (LUDW-) LUDWIG INST CANCER RES.  
 PI Chen Y, Gout I, Gure A, Ohare M, Odata Y, Old LJ,  
 PI Preundschuh M, Sahin U, Scanlan MD, Stockert E,  
 PI Tureci O;  
 DR WPI: 99-132448/11.  
 PT New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 PS Claim 67; Page 662; 787BP; English.  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 SO Sequence 1306 BP; 348 A; 331 C; 375 G; 252 T;

Query Match 58.4%; Score 1306; DB 1; Length 1306;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 931 aaaaatagccgagccttgacatctcatgttagctgagcagcgtggcgagcgtgttcacg 990  
 DB 1 AAAAATAGCCGCGAGCCTTGACATCTCATTTGATGTGACGTGGCGGAGCTGTTCATG 60

QY	991	acgagccgggagcgctcgcgagagcgcgcgagcgtgagctgagctcgcgagagagctctc	1050
Db	61	ACAGACCGGGAGCGGCTGGCGAGAGGGCGGGCACCGTGAGCTCACGGCAGAGAGTTGTC	120
QY	1051	atgcaagaagcggctcgcgagatgagagatccaaagaatccctcagagcagcagagaatgag	1110
Db	121	ATGCAGAAAGCGGCTGGCGATGAGATCCACMACAGATCTCTCAGAGACGACGAGAGATGGAG	180
QY	1111	cgcgcaagagagaaagaatctgcccagaagcgcgagcagagagaatgagatlacggaag	1170
Db	181	CGCGAAAGGAGAAATAATGGCCAGAAAGCGCAGCAGAGAAATAATGAGATACCGGAG	240
QY	1171	gagatgagacagatctgtagagagagagaagaattaaagcaatggygaagaagtctg	1230
Db	241	GAGATGGAACGATTTGTAGAGGAGGAAGAACTTTAACMACCAATGGGAAGAGACTGG	300
QY	1231	ggcccaagaagagacagcttaccctctgcttaaacataactgcgtgagatgtaacccagtaacc	1290
Db	301	GGCTCAAAAGAGACAGCTACTTGGCTTAACCATACTGCTGAGATACACCCAGTATACC	360
QY	1291	cttcgcaagcccaagatgtagcagggagtggaacctgagctcgagcccgacagatgacctg	1350
Db	361	CTTGCAAGGCCAAAGTATGATCAGGGGAGTGGAACTGAGACTGAGACCCGAGATGACCTG	420
QY	1351	gatgagagcacgagagcgagagcagagacagatctccggaatatgagaaagcttgagc	1410
Db	421	GATGAGAGCACGGAGGAGCAGGGAGAGAGAGATTTCCGAAATATGAGAAAGGCTTTGAC	480
QY	1411	ccctaccctctgttcaccccgagagagatcatggygaagagatgctccgctcctaagatc	1470
Db	481	CCCTACTCTATATGTCACCCAGAGACAGATCATGGGGAAGATGTCGGCTCTCTAGCAATC	540
QY	1471	aagaagagagagatctctagaacctggccctggagaggggtggtgagctccccaatgggaag	1530
Db	541	AAGAAGAGAGGATCTTATAGACTGCGCCCTGGAAGCGGCTGTGAGATCCCCATTGGAGAG	600
QY	1531	gtggtcgcttcctgctgtgtagtagcggggagagctgctgtagcggcatggtgtagatgtagaa	1590
Db	601	GTGGTCGTTCTGCTGTGTATGAGACGGGGAGCTGCTGACGGGCATGNGGCAATTGTGAA	660
QY	1591	gggagacgagatcatgagcaatccaacggcgaagattgtgacagctataccctggtcgagagct	1650
Db	661	GGGAGACGAGATCAAGGCAATCAACGGCAGATGTATACAGACTACCCCTGGCTGAGGCT	720
QY	1651	gagcgtccctcgcaagaagcgctggaatccaagggcgggagatgtagatcgacctgtgtgtcc	1710
Db	721	GACGCTGCCCTGCGAAGAGGCTCTGGAATCAGGGCGGGGATCTGATGACTTGTGTGTCC	780
QY	1711	gtctgccccccaagaagtagtagcagatgagcttctgtctgtagatccaagaagggga	1770
Db	781	GTCGCCCCCAAGGAGATATGACGATGAGACTGACTCTTGTCTGATGTCMAAAGGGGA	840
QY	1771	aaccaaattcacgcggttagaagaacagtgagctccggccccacctcgtagaacaagaagct	1830
Db	841	AACCAAAATTCACGGGTTAGGAAACAGTAGACTCCGGCCACCACTGTAACACAAAGGCT	900
QY	1831	cgagcagcgcttgagagagcgcaatgagacacacagagatgcatccttggaagcttgat	1890
Db	901	CGGACACGACTTGGAGAGAGCCCAATGACACACCAAGATGCAATCTTGGACCTGAT	960
QY	1891	ctataccacaggaatctccaactccctcttgccctgtaaaccaagggcgagatagaagaagc	1950
Db	961	CTATACCCACGAGATCTCAAAATCCCTTGGGCCCTAACCAGAGCCAGATATAGAGAACGC	1020
QY	1951	tcggggcaatttttgaagcacaatgtgagagaagggagcgagccgcttgaggaga	2010
Db	1021	TCGGGCCACTTTTTTGAAGGCCAATGTGGAGGAAAGGAGCAGCCAGCCGTTTGGAGAA	1080
QY	2011	gattcacaagatccaagactcaatctctctccctcgcgccaatggaattgtgtctcca	2070
Db	1081	GATCTCAAGATCCAGACTCTCATTTCTTTCTCTCTGCCCCAGATGATTTTGGTCTCTCCA	1140
QY	2071	gctttggggagctcctctccttgagaccttaagaagcccaatgagagctctctcctca	2130

Dd	1141	GCCTTGGGAGCATCCTTCCTTGACCCCTTAATGAAGACCACACGAGGATCTCTCTGCCA	1200
Oy	2131	tcccttcctcgcacctcgctctaattgctgcgaagatgtgcactccaacattacct	2190
Dd	1201	TCCCTCTCTCTCGCCCTCTGCTCTTAATTCTGTCCAGAGATTGTCATCCAAACTTACTCT	1260
Oy	2191	gagctcatataataaacagatttatttcacagttaaaaaa	2236
Dd	1261	GAGCTCATTAATAATAACAGATTATTTCAGCTTAATAAAA	1306
 RESULT 7 V73805/c ID V73805 standard; DNA: 32207 BP. AC V73805; DT 25-FEB-1999 (first entry) DE KSHV LUR DNA (nucleotides 105,301-137,507). KW Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2; KM dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis; KN diagnosis; treatment; HHV8; capsid protein IV; tegument protein IV; KV glycoprotein; kaposin; cyclin D; immediate early protein; IEP; OX-2; KM v-abl; G-protein coupled receptor; FCARAT; ds. ON Kaposi's sarcoma-associated herpesvirus. PS US5849564-A. PD 15-DEC-1998: PF 29-NOV-1996: 770379. PR 29-NOV-1996: US-770379. PA (UYCO) UNIV COLUMBIA NEW YORK. PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ; DR Kaposi's sarcoma-associated herpes virus nucleic acid - encodes PT or dihydrofolate reductase and is useful for treatment, prophylaxis PR or diagnosis of Kaposi's sarcoma PS Disclosure: Column 155-182; 109pp: English. This sequence is a fragment of the Kaposi's sarcoma-associated herpesvirus (KSHV) LUR (long unique region). This fragement contains coding regions for ORF65 which encodes capsid protein IV, ORF66, ORF67 which encodes tegument protein IV, ORF68 which encodes a glycoprotein, ORF69, K12 which encodes kaposin, K13, ORF72 which encodes cyclin D, ORF73 which encodes immediate early protein (IEP), K14 which encodes OX-2 (v-abl), ORF74 which encodes G-protein coupled receptor, ORF75 which encodes tegument protein/FCARAT, K15. KSHV is a new human Herpesvirus (HHV) believed to cause Kaposi's sarcoma (KS) which is the most common form of neoplasm occurring in persons with acquired immune deficiency syndrome (AIDS). The DHFR protein is useful for vaccination, prophylaxis, diagnosis and treatment of a subject with Kaposi's sarcoma and for detecting expression of a DNA virus associated with Kaposi's sarcoma in a cell.			
SQ	Sequence	32207 BP;	7229 A; 9156 C; 8713 G; 7109 T;
 Query Match 3.1%; Score 68.4; DB 1; Length 32207; Best Local Similarity 47.6%; Pred. No. 9.6e-08; Matches 201; Conservative 0; Mismatches 221; Indels 0; Gaps 0;			
Oy	979	gagctgttatgaaccaggaccggaggggctgcgcaagggcgcgagcagcytgagctgcaagg	1038
Dd	19815	GAGCAGCACACACAGATGATGACGACGACGACGAGATGACACACGACGATGATGACGAG	19756
Oy	1039	caggagcttcctcatgagaagaagcggtctgtagatggagatccacaagaatcctccaggagca	1098
Dd	19755	CAGCAGCAGAGATGAMCAGACGACGACGACGACGACGACGACGACGACGACGACGACGAG	19656
Oy	1099	caggagatgagcagcggaagaagaagaatgtcccagaagcgacagaggaanaatgag	1158
Dd	19695	GAGTTTAGAGAGCAGACGACGACGAGATTACGAGATCAGAGCAGAGCACTTGAGAGCAGAG	19636
Oy	1159	agatccggaagaagatggaacagatttgaaaggaggggaagagaagtcttaagaagaacaggy	1218
Dd	19635	CAGGAGTTTAGAGAGCAGAGCAGGAGTTTAGAGGAGCAGAGACGACGAGACTTACAGGAGCAG	19576
Oy	1219	gaagaagaactctggggctccaagaagaacagctactccttgctcaaaaacatcacttctaagta	1278

DB 19575 GACGACGAGATTAGAGGAGCAGGAGCAGACTTACAGAGCAGAGCAGATTAGAGAG 19516  
QY 1279 caccacgtacccttcgcagacccaagtatgacagggatggaacctgagccgcagccc 1338  
DB 19515 CAGGACGACGAGATTAGAGAGCAGGAGCAGACTTACAGAGCAGAGGTGGAAGAGCAA 19456  
QY 1339 gcagatgacctgagtgaagcaggaagcaggaagcagagatctccgaaataatga 1398  
DB 19455 GAGCAGAGAGGTGAGAGCAGAGCAGAGCAGAGAGCAGAGATTAGAGAGGTGGAG 19396  
QY 1399 ga 1400  
DB 19395 GA 19394  
RESULT 8  
ID V19941/c  
AC V19941 standard; DNA; 137507 BP.  
DE 03-AUG-1998 (first entry)  
KW KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;  
KW Interleukin-6; IL-6; Interferon regulatory factor; Rheumatoid arthritis;  
KW complement-binding protein; glycoprotein; capsid protein IV; infection;  
KW lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;  
KW HIV immune status; anti-inflammatory agent; therapy; ds.  
OS Kaposi's sarcoma-associated herpes virus.  
FH Key Location/Qualifiers  
FT CDS 1142..2794  
FT /tag= a  
FT /product= complement-binding protein  
FT 8699..11236  
FT /tag= b  
FT /product= glycoprotein B  
FT complement (11761..11875)  
FT /tag= c  
FT /product= Interleukin 6  
FT complement (21548..21832)  
FT /tag= d  
FT /product= macrophage inflammatory protein II  
FT complement (27137..27424)  
FT /tag= e  
FT /product= interferon regulatory factor 1  
FT 28661..29741  
FT /tag= f  
FT /product= protein T1.1  
FT complement (58976..60175)  
FT /tag= g  
FT /product= glycoprotein M  
FT complement (69412..69915)  
FT /tag= h  
FT /product= glycoprotein L  
FT complement (88410..88910)  
FT /tag= i  
FT /product= interferon regulatory factor 2  
FT 89600..90541  
FT /tag= j  
FT /product= interferon regulatory factor 3  
FT 90173..90643  
FT /tag= k  
FT /product= glycoprotein X  
FT complement (93636..94127)  
FT /tag= l  
FT /product= interferon regulatory factor 4  
FT complement (111931..112443)  
FT /tag= m  
FT /product= capsid protein IV  
FT complement (123808..127296)  
FT /tag= n  
FT /product= immediate early protein  
FN W09804576-A1.

PD 05-FEB-1998.  
PF 22-JUL-1997; U13346.  
PR 29-NOV-1996; US-757669.  
PR 25-JUL-1996; US-686243.  
PR 25-JUL-1996; US-686349.  
PR 25-JUL-1996; US-686350.  
PR 25-JUL-1996; US-687253.  
PR 25-JUL-1996; US-688814.  
PR 25-SEP-1996; US-708678.  
PR 10-OCT-1996; US-728323.  
PR 13-NOV-1996; US-747887.  
PR 13-NOV-1996; US-748640.  
PA (UYCO ) UNIV COLUMBIA NEW YORK.  
PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;  
DR WPI: 98-130615/12.  
PT New nucleic acid encoding Kaposi's sarcoma associated herpes virus  
PT proteins - useful for, e.g. detecting levels of HHV8 in, and  
PT preparation of vaccines for treatment of, HIV patients  
PS Example 2; Page 135-203; 230pp; English.  
CC This sequence represents the long unique region and terminal repeat of  
CC the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known  
CC as human herpes virus 8 (HHV8). This sequence contains the DNAs of the  
CC invention which encode KSHV polypeptides selected from: (a) viral  
CC macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);  
CC (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;  
CC (d) capsid protein IV encoded by ORF65; and (e) immediate early protein  
CC encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded  
CC by it, and antibodies (Ab) specific for the proteins are useful for  
CC detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body  
CC fluids or tissue samples. HHV8 infections can be treated with antisease  
CC or triplex forming molecules or agents that bind specifically to the  
CC protein. Ab may be used for prophylaxis or treatment of HHV8 infection,  
CC while the protein can be used in protective vaccines. Ab may also be used  
CC to differentiate between lymphomas, and HHV8 may be implicated in many  
CC other lymphoproliferative diseases such as lymphomas, leukaemia,  
CC splenomegaly and mycosis fungoides. Cells and animals containing the  
CC nucleic acid are useful for drug screening. HHV8-derived peptides can be  
CC used as targets for antiviral drugs, e.g. dihydrofolate reductase gene  
CC can be inhibited with methotrexate. These can also be used to determine  
CC the immune status of a patient infected with HIV. HHV8 derived protein  
CC viral MIP III may be used as an anti-inflammatory agent for,  
CC e.g. treating Rheumatoid arthritis. This sequence is stated as containing  
CC 81 open reading frames.  
SQ Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T;  
Query Match 3.1%; Score 68.4; DB 1; Length 137507;  
Best Local Similarity 47.6%; Pred. No. 1.9e-07;  
Matches 201; Conservative 0; Mismatches 221; Indels 0; Gaps 0;  
QY 979 gagctgttcattacacagccggagcgctgcgcaggaagcgcgagcgctgagctgcagcg 1038  
DB 125115 GAGCAGCAGCAGCAGCAGATGATGACGACGACGAGATGACGACGACGATGACGAG 125056  
QY 1039 caggagcttcattcattcagaagcgctgcgcagtgagtgacccaagaagctcctcagaagcag 1098  
DB 125055 CAGCAGCAGGATGATGACGAGCAGCAGGAGGAGCAGACAGCAGAGGAGCAGCAGCAG 124996  
QY 1099 caggagatgagcgcggaaggaagaaatgtccccaagcgagcagcaaggaagaaatgag 1158  
DB 124995 GAGTTAGAGGAGCAGGAGCAGGAGTTAGAGATCAGGACGACGAGTGTAGAGGAGGAG 124936  
QY 1159 agataccggaaggaagatggaacagatgtcagaggaaggaagaagttaagaagcaatg 1218  
DB 124935 CAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGGAGCAGGACGACGAGATGAGAGCAG 124876  
QY 1219 gaagaagactggggctcaagaagaaagctactcttgcctaaacaatcactcgtctgagta 1278  
DB 124875 GAGCAGAGATTAGAGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGCAGGAGTTAGAGAG 124816  
QY 1279 caccacgtacccttcgcagacccaagtatgacagggatggaacctgagccgcagccc 1338  
DB 124815 CAGGACGACGAGATTAGAGAGCAGGAGCAGGAGTTAGAGGAGCAGGAGTTGGAAGAGCAA 124756

Oy	1339	gcagatgaccctggagtgtgagagcaggaagagagagagacagatcttcgcggaatatgag	1398
Db	124755	GAGCAGGAGGTGTGAAAGACCAAGCAGCAGGACGAGAAAGCAGCAATTAGAGAGGTGGAC	124656
Oy	1399	ga	1400
Db	124695	GA	124694

Accession	Gene	Protein	Score	DB	Length	Query Match
V55831	9		2.6%	Score 58;	DB 1;	Length 799;
ID	V55831	standard; DNA: 799 BP.	Best Local Similarity	46.9%;	Pred. No. 8.3e-06;	
AC	V55831:		Matches 181;	Conservative 0;	Mismatches 205;	Indels 0; Gaps 0
DT	18-NOV-1998	(first entry)				
DE	Nucleotide sequence of the stabilising polypeptide; proteolytic degradation.					
FW	Fusion protein; stabilising polypeptide; proteolytic degradation.					
KW	resistance: half-life; autoimmune disease; inflammation; nitro drug;					
KW	IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;					
KW	nitroreductase protein; enzyme therapy; prodrug therapy; protease;					
OS	cancer; pathological condition; ss.					
SV	Epstein-barr virus.					
PN	MO9822577-AL.					
PD	28-MAY-1998.					
PF	17-NOV-1997.	IB1508.				
PR	25-JUN-1997;	US-048945.				
PR	15-NOV-1996;	US-030986.				
PA	(MASUCI) MASUCCI M G.					
PI	Masucci MG;					
DR	WPI: 98-312463/27.					
PT	New fusion proteins resistant to proteolytic degradation -					
PT	comprising a core protein with a stabilising polypeptide comprising					
PT	a peptide sequence containing glycine repeats					
PS	Disclosure: Fig 4B: 120pp: English.					
CC	This is a nucleotide sequence of the stabilising sequence-encoding					
CC	insert. The invention provides a method for increasing the resistance					
CC	of a core protein to proteolytic degradation that comprises linking or					
CC	inserting onto or into the core protein a stabilising polypeptide of					
CC	formula [(Gly)nX(Gly)nY(Gly)nZ)n where Glya, Glyb, Glyc are 1-6					
CC	sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,					
CC	Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not					
CC	be identical from n repeat to n repeat. Alternatively a nucleic acid					
CC	encoding the stabilising polypeptide can be linked onto or inserted into					
CC	a nucleic acid encoding a core protein. The fusion proteins of the					
CC	invention are more resistant to degradation by proteases and, thus, have					
CC	a longer half-life than the unfused core protein. The products can be					
CC	used for treating autoimmune diseases, cancer and inflammation. In					
CC	particular, the core protein may be an IkappaB regulator protein for the					
CC	treatment of inflammatory bowel disease, or a nitroreductase protein					
CC	which can activate nitro drugs in enzyme/prodrug therapy to treat cancer					
CC	or other pathological conditions. The fusion proteins can also be used in					
CC	diagnostic methods such as in vivo imaging.					
SO	Sequence 799 BP;	201 A;	106 C;	479 G;	13 T;	

Db	525	GAGGGGCGAGGAGGGCGAGGAGGAGGGGCGAGGAGGAGCGAGGAGGCGAG	58b
Qy	1237	aaggaacgctactcttctcctaataacatcaactctgtaggtacaccagtaaccttcgc	1237
Db	585	GAGGGGCGAGGAGCGAGGAGGAGGAGGAGGAGGCGAGGAGCGAGGAGGCGAG	644
Qy	1297	aagccaagaatgatgcagagggagtggaaactagctcgcagcccgcaagatgcactgtatgga	1337
Db	645	GAGCGAGGAGGGGCGAGGAGGCGAGGAGGAGGAGGAGGAGGCGAGGAGGAGGAG	700
Qy	1357	ggcacggagagacgagaggaagcagga	1382
Db	705	GCGCAGGAGCGAGGAGGCGAGGAGCA	730
RESULT 10			
ID	V21683		
AC	V21683	standard; DNA; 9600 BP.	
DT	17-AUG-1998	(first entry)	
DE	Vector plasmid PCMVKmTr-Epi.		
KM	Polynucleotide delivery; Plasmid PCMVKmTr-Epi; vector;		
KW	gene therapy; vaccine; polycationic agent; ss.		
OS	Chimeric - Epstein-Barr virus.		
OS	Chimeric - Adeno-associated virus.		
OS	Chimeric - Cytomegalovirus.		
OS	Chimeric - Bos taurus.		
FH	Key	Location/Qualifiers	
FH	CDS	14..2594	
FT		/*tag= a	
FT		/product= "EBV nuclear antigen A"	
FT	misc_feature	2623..4559	
FT		/*tag= b	
FT		/note= "EBV origin of replication"	
FT	repeat_unit	4928..5104	
FT		/*tag= c	
FT		/rpt_type= INVERTED	
FT		/note= "AAV Inverted terminal repeat"	
FT	repeat_unit	7189..7355	
FT		/*tag= d	
FT		/rpt_type= INVERTED	
FT		/note= "AAV Inverted terminal repeat"	
FT	promoter	5112..6734	
FT		/*tag= e	
FT		/note= "CMV immediate-early enhancer/promoter"	
FT	terminator	6818..7050	
FT		/*tag= f	
FT		/note= "bovine growth hormone polyA sequence"	
PN	W09806437.A2.		
PD	19-FEB-1998.		
PF	13-AUG-1997; U14465.		
PR	13-AUG-1996; US-023867.		
PA	(CHR ) CHIRON CORP.		
PI	Cohen F, Dubois-Stringfellow N, Dworki V, Innis MA,		
PI	Murphy JE, Tetsuo U, Zukermann R;		
DR	WPI; 98-159296/14.		
PT	Polycationic agents based on alpha-amino acids, able to complex		
PT	with nucleic acid - to facilitate its entry into cell, condense it		
PT	and protect it against serum degradation, particularly for use in		
PT	gene therapy		
PS	Disclosure; Page 77-80; 100pp; English.		
CC	This polynucleotide comprises the DNA sequence of vector plasmid		
CC	PCMVKmTr-Epi, which contains an Epstein-Barr virus (EBV) origin		
CC	of replication from plasmid pCEP4, a coding region for EBV nuclear		
CC	antigen A from pCEP4, a pair of inverted terminal repeats from		
CC	adeno-associated virus, a cytomegalovirus enhancer/promoter, a		
CC	bovine growth hormone polyA sequence, and a kanamycin resistance		
CC	selectable marker. Polynucleotides encoding polypeptides, such as		
CC	erythropoietin or leptin, and ribozymes and antisense		
CC	polynucleotides can be inserted into the vector. The vector is		
CC	preferred for use in novel compositions and methods for improved		
CC	polynucleotide delivery into cells. In these methods, polycationic		
CC	agents are used to increase the frequency of uptake of a		

CC nucleic acid (see also V21684-86) into a cell. The polycationic  
CC agent can condense with the nucleic acid and inhibit serum and/or  
CC nuclease degradation of the nucleic acid. The nucleic acid can be  
CC a vector, may express a therapeutic protein or a vaccinating viral  
CC or cancer antigen, or is itself therapeutic (antitense or  
CC ribozyme). The methods and compositions can be used in the gene  
CC therapy of many diseases.  
SQ Sequence 9600 BP; 2336 A; 2376 C; 2817 G; 2081 T;

Query Match	2.68;	Score 58;	DB 1;	Length 9600;
Best Local Similarity	46.98;	Pred. No. 2.8e-05;		
Matches 181; Conservative	0;	Mismatches 205;	Indels 0;	Gaps 0;

[illegible]

RESULT	11
051731	
ID	Q51731 standard; DNA; 10596 bp.
AC	Q51731;
DT	31-MAY-1994 (first entry)
DE	Plasmid pCISBON for subcloning huHGF variants.
KM	Hepatocyte Growth Factor; HGF; variant; mutein; in vitro mutagenesis;
KW	Proteolysis resistant; liver; malignancy; CMV-driven;
KM	Cytomegalovirus; episomal expression plasmid; ss.
OS	Synthetic.
FH	Key
FT	enhancer
FT	1..611
FT	Location/Qualifiers
FT	/*tag= a
FT	/note= "CMV enhancer/promoter"
FT	758..775
FT	promoter
FT	/*tag= b
FT	/label= SP6-promoter
FT	845..849
FT	misc-feature
FT	/*tag= c
FT	/note= "SP6 RNA start"
FT	902..966
FT	misc-feature
FT	/*tag= d
FT	/function= cloning_linker
FT	967..1107
FT	polya-signal
FT	/*tag= e
FT	/note= "SV40 poly A"
FT	1108..1531
FT	misc-feature
FT	/*tag= f
FT	/function= SV40_origin
FT	1580..4189
FT	misc-feature

```

FT      /tag= g
FT      /label= EBNA-1
FT      4190. .6374
FT      /tag= h
FT      /function= orip
FT      4295. .4887
FT      /tag= i
FT      /note= "family of repeats"
FT      5866. .5978
FT      /tag= j
FT      /note= "dyad region"
FT      6375. .6457
FT      /tag= k
FT      /label= HSV-TK_terminator_3'-end
FT      6975. .7975
FT      /tag= l
FT      /phenotype= neomycin_resistance
FT      /note= "rns neomycin phosphotransferase gene"
FT      7975. .8112
FT      /tag= m
FT      /label= TK_promoter
FT      8114. .8594
FT      /tag= n
FT      /function= M13_ori
FT      8595. .10414
FT      /tag= o
FT      /label= delta_2a
FT
FT      WO9323541-A.
FT      25-NOV-1993.
FT      17-MAY-1993; US-004648.
FT      18-MAY-1992; US-884811.
FT      18-MAY-1992; US-885971.
FT      (GETH ) GENENTECH INC.
FT      Godowski PJ, Lokker NA, Mark MR;
FT      WPI; 93-386573/48.
FT      DR
FT      Hepatocyte growth factor variants - are resistant to proteolytic
FT      cleavage into its two-chain form, used to treat malignancies
FT
FT      associated with HGF receptor
FT
FT      Example 1; Fig 6; 87pp: English.
FT      Plasmid pcisEBON (a pRK5 derivative) is an episomal CMV driven
FT      expression plasmid. HuHGF variants with enhanced receptor binding
FT      activity were produced by site-directed mutagenesis. Stable
FT      populations of preferred HGF variants were obtained by transfecting
FT      human embryonic kidney 293 cells and then these were subcloned in
FT      pcisEBON. See R52940-R52949 for examples of pref. HGF variants.
FT      Sequence 10596 BP; 2625 A; 2571 C; 3024 G; 2376 T;

```

Query Match	2.6%	Score 58	DB 1	Length 10596
Best Local Similarity	46.9%	Pred. No. 3e-05		
Matches 181	Conservative 0	Mismatches 205	Indels 0	Gaps 0
QY 997	cggagacgcgtgcagaaagcgcgcagcgttaagctgcagcgcgcaggaacttcatactcag	1056		
Db 2550	CAGGACGAGGAGGAGGCGCAGAGCAGGAGGGGCGAGAGGGGCGAGAGGGCGAGCAG	2609		
QY 1057	aagcgcgtgcgatgttagctccacaagaatcctccagagcaagcagagatgacgcgcga	1116		
Db 2610	GAGGGCGAGGAGCAGGAGGAGGCGGCGCAGGAGGGGCGCAGGAGCAGGAGGGCGAG	2669		
QY 1117	aggagaaaaagaattgcccaagagcagcagagagaaaaatgagatataccggaagagatg	1176		
Db 2670	GAGCAGGAGGGGCGAGGACAGGAGGAGGCGGCGCAGAGCAGAGGGGCGAGGCGGCGAG	2729		
QY 1177	gaacagattgtagagagagagaagaagtttaagaagcaatgaggaaagaagtgggtcca	1236		
Db 2730	GAGGGCGAGGAGGGGCGAGGACGAGGAGGGGCGCAGGAGGAGGAGGAGGGCGAG	2789		
QY 1237	aaggaacgcctactctcttgccataaacatcaactgcttgaggtacaccacgaatacccttcgc	1296		
Db 2790	GAGGGCGAGGACGAGGAGGCGCAGGAGGGGCGCAGAGGAGGCGCAGGAGGGCGAG	2849		
QY 1297	aagccaagatlaactcagagatgaaacctgaactgcgcagcccgacatctacctgatatga	1356		

QY	1357	gacacgagagagcagagagagagagagga	1382
Db	2850	GAGCAGAGAGGGGACAGAGGGCGACGACGAGGGCGACAGAGGGCGACGAGAGGAG	2909
QY	1357	gacacgagagagcagagagagagagagga	1382
Db	2910	GGCGAGAGAGCAGAGAGGGCGACGAGCA	2935
RESULT	12		
T40348			
ID	T40348	standard; DNA; 10596 BP.	
DT	T40348:		
DE	09-DEC-1996	(first entry)	
KW	Plasmin	plcSEBON for expression of hepatocyte growth factor.	
KW	Human; hepatocyte growth factor; HGF; huHGF; serum; proteolytic cleavage;		
KW	pro-hormone; beta subunit; alpha subunit; kringle domain; prothrombin;		
KW	plasminogen; catalytic domain; serine protease; HGF variant;		
OS	HGF receptor; malignancy; chronic HGF receptor activation; ss.		
PN	Synthetic		
PF	US5547856-A.		
PR	20-AUG-1996		
PR	18-MAY-1992; 884811.		
PR	18-MAY-1992; US-885971.		
PR	18-MAY-1992; US-884811.		
PR	13-JUL-1993; US-087783.		
PI	(GETH ) GENENTECH INC.		
PI	Godowski PJ, Lorker NA, Mark MR;		
PI	WPI, 96-392634/39.		
PT	New hepatocyte growth factor variants - are resistant to in vivo		
PT	proteolytic cleavage into a 2-chain form, useful as HGF antagonists		
PS	Example 1; Fig 6; 39pp; English.		
CC	This sequence represents the episomal CMV driven expression plasmid		
CC	plcSEBON which was used in the expression of variant human hepatocyte		
CC	growth factor (HGF). HGF is isolated from human serum and is a		
CC	disulphide linked heterodimer derived by proteolytic cleavage of the		
CC	pro-hormone between residues 494 and 495. This generates a molecule		
CC	composed of an alpha subunit of 440 amino acids (mol. wt. 69 kD) and		
CC	a beta subunit of 234 amino acids (mol. wt. 34 kD). The alpha and beta		
CC	subunits are encoded by a single open reading frame. The alpha subunit		
CC	contains four kringle domains based on their homology to kringle-like		
CC	domains in other proteins, e.g. prothrombin, plasminogen. The beta		
CC	subunit shows high homology to the catalytic domain of serine proteases.		
CC	However two of the three residues which form the catalytic triad of		
CC	serine proteases are not conserved in HGF. Therefore, the precise		
CC	function of the beta chain remains unknown. The invention includes HGF		
CC	variants which retain HGF receptor binding activity without having the		
CC	biological activity of wild-type HGF. They can be used for the treatment		
CC	of pathological conditions associated with the activation of a HGF		
CC	receptor such as malignancies associated with chronic HGF receptor		
CC	activation. The plcSEBON plasmid comprising the HGF coding sequence		
CC	may be used for manipulation of the HGF coding sequence and expression		
CC	of the variant HGF's of the invention.		
CC	Sequence 10596 BP; 2627 A; 2571 C; 3023 G; 2375 T;		
Query Match	2.6%;	Score 58;	DB 1; Length 10596;
Best Local Similarity	46.9%;	Pred. No. 3e-05;	
Matches 181;	Conservative	0;	Mismatches 205; Indels 0; Gaps 0;
QY	997	cgagagcggctgacagagcgcgacgctgagctgacgcgacgagagcttcacgcaag	1056
Db	2550	CAGGACGACGAGAGAGGGCGACGAGCGAGGCGGACGAGGCGGACGAGGCGACGAGCGAG	2609
QY	1057	aagcggtgctgcatgtaagtcacaagaatccctccaggaacacagagaagtatgagcgca	1116
Db	2610	GAGGGCGACGACGACGAGGAGGAGGCGCGAGGCGGACGAGGCGGACGAGGAGGCGCGAG	2669
QY	1117	agagagaaagaatctccagaagagcaagagaggaagaaatgagagatacccggaagagatg	1176
Db	2670	GAGCAGAGAGGGCGACGAGCGAGGCGGACGAGCGAGGCGGACGAGGCGGACGAGCGAG	2729
QY	1177	gaacagatgttagagagagagaagaatttaagaagcaatgggaagaaagactggggctca	1236

DB	2730	GAGGGGCGAGGAGGGGCGACGACGAGGAGGGGCGACGAGGGGCGAGGAGGAGGAGGGGCGAG	2789			
QY	1237	aaggaacgactactcttgccttaaaccaactactgctgtgagtaaccagttacccttcgc	1296			
DB	2730	GAGGGGCGAGGAGGAGGGGCGACGAGGAGGGGCGACGAGGAGGAGGGGCGAG	2849			
QY	1297	aagccaagatgatcatcagggagtgagtaactgtgcttcgagccgcagatgactgtatgta	1356			
DB	2850	GAGCGAGGAGGGGCGACGAGGGCGACGAGGAGGGGCGACGAGGGGCGAGGAGGAGG	2909			
QY	1357	ggcagggagagcaggagagcagga	1382			
DB	2910	GGGCGAGGAGGAGGAGGGGCGACGAGCA	2935			
RESULT 13						
ID	X15650					
AC	X15650:	standard; DNA; 10596 BP.				
DT	10-MAY-1999	(first entry)				
DE	Nucleotide sequence of plasmid pCIS.EBON.					
KW	Plasmid pCIS.EBON; hepatocyte growth factor; HGF; variant;					
KM	HGF receptor activation; ss.					
OS	Synthetic.					
PN	U5879910-A.					
PD	09-MAR-1999.					
PF	09-FEB-1994; 194087.					
PR	09-FEB-1994; US-194087.					
PA	(GERTH ) GENENTECH INC.					
PI	Godowski PJ, Loker NA, Mark MR;					
PT	WPI; 99-203949/11.					
PR	New hepatocyte growth factor variants - useful for studying					
PT	structure-function relationships in the wild type molecule and for					
PT	treating conditions associated with chronic hepatocyte growth factor					
PT	receptor activation					
PS	Example 1: Fig 6A-F; 40pp; English.					
CC	The present sequence represents the nucleotide sequence of pCIS.EBON,					
CC	which is used in the course of the invention. The specification					
CC	describes a hepatocyte growth factor (HGF) variant (HGfV) comprising an					
CC	amino acid (aa) alteration at or adjacent to position 692 of the					
CC	wild-type human HGF (huHGF) aa sequence. HGfV may be used in studies to					
CC	identify the structure-activity relationships of HGF in order to identify					
CC	the functionally important domains in the aa sequence. It may also be					
CC	used to identify aa residues which are responsible for the interaction of					
CC	HGF with its receptor, and those aa that are responsible for the					
CC	biological activity of it. Variants of HGfV which have enhanced receptor					
CC	binding affinity (compared to wild-type huHGF) and are more biologically					
CC	active than wild type huHGF, may be used as huHGF agonists. Conversely,					
CC	variants of HGfV which have enhanced receptor binding affinity (compared					
CC	to wild-type huHGF) but which are biologically inactive may be used as					
CC	huHGF antagonists, and may be used to block the binding of wild-type					
CC	huHGF to its receptor. This permits the treatment of pathogenic					
CC	conditions associated with the activation of an HGF receptor, such as					
CC	malignancies associated with chronic HGF receptor activation.					
CC	HGfV-immunoglobulin chimeras may be produced (by standard methods) and					
CC	used in protein A purification, immunohistochemistry, and					
CC	immunoprecipitation techniques (in place of anti-HGF antibodies) or in					
CC	screening studies to identify inhibitors of HGF-HGF interactions.					
SQ	Sequence 10596 BP; 2627 A; 2571 C; 3023 G; 2375 T;					
Query Match 2.6%; Score 58; DB 1; Length 10596;						
Best Local Similarity 46.9%; Pred. No. 3e-05;						
Matches 181; Conservative 0; Mismatches 205; Indels 0; Gaps 0						
QY	997	cggagcggcttgcgcagagcgcgcgacgctgagctgcagcggcagagagcttcctatgcag	1056			
DB	2550	CAGGAGCAGAGAGAGAGGCGACGAGCAGAGGAGGCGACGAGGAGGCGACGAGCAG	2609			
QY	1057	aagcggtctgcgatactgagatcccaaatctctccagggagcagcaggagatgtgagcgcaa	1116			
DB	2610	GAGGGGCGACGAGCAGAGGAGGGGCGACGAGGGGCGAGGAGGGGCGACGAGGAGGCGAG	2669			



QY	1117	agcgaaaaataattgccccaagcgcaagcaaaaatctgagatctccggaagaagatg	1176
Db	2670	GAGCAGAGGGGGCGAGACAGAGAGGGGCGAGACAGAGAGGGCGAGAGGGCGAGAGCAG	2729
QY	1177	gaacagatctgtagaagagaagaagaagcttaagaagcaatggygaagaacttgygtccta	1236
Db	2730	GAGGGCGAGAGGGGCGAGAGCGAGAGGGGCGAGAGAGGCGACAGAGAGAGAGGGCGAG	2789
QY	1237	aaggaacagctactcttgccttaaacatcactgcctgagatgtaacccactaccctctgc	1286
Db	2790	GAGGGGCGAGACGAGAGAGAGGGGCGAGAGGGGCGAGACAGAGAGGGCGAGAGGGCGAG	2849
QY	1297	aagccaaagatgatcatcgagagctgtgaaccttgatctgaagcccgacagatgatctgtatga	1356
Db	2850	GAGCAGAGAGGGGCGAGAGGGGCGAGAGAGAGGGCGAGAGAGGGCGAGAGAGAGAGAG	2909
QY	1357	ggcagcgagagcgacgagagagcgagca	1382
Db	2910	GGGCAGGAGCAGAGGGGCGAGAGCA	2935

RESULT 14	
ID	V55830
AC	V55830 standard; DNA; 795 BP.
DT	18-NOV-1998 (first entry)
DE	FLCA insert stabilising polypeptide encoding DNA.
KW	Fusion protein; stabilising polypeptide; proteolytic degradation; resistance; half-life; autoimmune disease; inflammation; nitro drug; IkappaB regulator protein; inflammatory bowel disease; in vivo imaging; nitroreductase protein; enzyme therapy; prodrgn therapy; protease; cancer; pathological condition; ss.
KM	Epstein-Barr virus.
OS	Location/Qualifiers
FT	Key
FT	1..788
FT	/tag- a
FT	/product- "stabilising polypeptide"
PN	W09822577-A1.
PD	28-MAY-1998.
PF	17-NOV-1997; IB1508.
PR	25-JUN-1997; US-048945.
PR	15-NOV-1996; US-030986.
PA	(MASU/) MASUCCI M G.
PI	Masnuci MG;
DR	WPI; 98-312463/27.
DR	P-PSDB; W79128.
PT	New fusion proteins resistant to proteolytic degradation - comprising a core protein with a stabilising polypeptide comprising a peptide sequence containing glycine repeats
PS	Discloser; Fig 3; 120pp; English.
CC	This DNA encodes a stabilising polypeptide and is the FLCA insert of the invention. The invention provides a method for increasing the resistance of a core protein to proteolytic degradation that comprises linking or inserting onto or into the core protein a stabilising polypeptide of formula (Glya)X(Glyb)Y(Glyc)Zn where Glya, Glyb Glyc are 1-6 sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met, Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not be identical from n repeat to n repeat. Alternatively a nucleic acid encoding the stabilising polypeptide can be linked onto or inserted into a nucleic acid encoding a core protein. The fusion proteins of the invention are more resistant to degradation by proteases and, thus, have a longer half-life than the unfused core protein. The products can be used for treating autoimmune diseases, cancer and inflammation. In particular, the core protein may be an IkappaB regulator protein for the treatment of inflammatory bowel disease, or a nitroreductase protein which can activate nitro drugs in enzyme/prodrgn therapy to treat cancer or other pathological conditions. The fusion proteins can also be used in diagnostic methods such as in vivo imaging.
CC	Sequence 795 BP; 200 A; 104 C; 478 G; 13 T;

Query Match	2.58;	Score 57;	DB 1;	Length 795;
Best Local Similarity	51.88;	Pred. No.	1.5e-05;	

Matches	129;	Conservative	0;	Mismatches	120;	Indels	0;	Gaps	0
QY	997	cgagagcgagctgycagagagcgcgcaagcgltgagctgycagcgagagcttctatcgag							1056
Db	342	CAGGACGACAGAGAGAGGCGCAGGACGAGAGGGGCAAGAGGGGCGCAGGAGGAC							401
QY	1057	aagcgagctbgcgatgtagtcccaacaagatctctccaggaagcagcagagagaatggaacgca							1116
Db	402	GAGGGGCGACGAGCAGGAGGAGGGGCGAGGGGCGACGAGGGGCGAGGAGGCGGCGAG							461
QY	1117	agggagaaaaaagaattgcccagaagcgacgacgaggggaaatagagagataccgggaagggatag							1176
Db	462	GAGCAGAGAGGGGCGACGAGGAGGGGCGAGGACGAGAGGGGCGAGGCGCGAGGACGAG							521
QY	1177	gaacagatgttagaggaaggaagagaagtttaagaaacaaatgggaagaaagagctggggctca							1236
Db	522	GAGGGGCGACGAGGAGGGGCGAGGAGGGGCGAGAGGGGCGACGAGGAGGAGGGGCGAG							581
QY	1237	aaggaacag	1245						
Db	582	GAGGGGCGAG	590						

RESULT 15

023092

ID 023092 standard; DNA; 543 BP.

AC 023092;

DE 17-AUG-1992 (first entry)

DE Antigen tc-7a gene.

KW Oocysts; Mab 12-07; sporozoite; ss.

OS Elmeria tenella.

PN M09204460-A.

PD 19-MAR-1992.

PF 05-SEP-1991; U06430.

PR 12-SEP-1990; U5-581693.

PA (GENE-) GENEX CORP.

PI Jacobson JW, Strausberg RL, Wilson SD, Pope SH, Strausberg SL; Ruff MD, Augustine PC, Danforth HD; Wpi: 92-114365/14.

DR P-P-SDB; R22392.

DR Vaccine against avian coccidiosis - comprising recombinant Elmeria antigen ac-1b or ac-6b gene, or microorganisms expressing them

PT Claim 5; Page 39 + Fig 5; 56pp; English.

PS To identify antigens of *E. tenella*, expression libraries were prepared in lambda vector, lambda gIII, using cDNA prepd. from polyA mRNA isolated from *E. tenella* oocysts. The cDNA expression library was screened with monoclonal antibody (Mab) 12-07 which was raised against the sporozoite stage of *E. tenella*. The library to be screened was plated on a host that allows lysis and plaque formation. During induction of the antigens encoded by the phage, the plaques were identified by screening the filters with Mab 12-07

CC The cDNA inserts from the Mab 12-07 positive phage were cloned into bacteriophage M13 and subjected to sequence analysis. Following sequence analysis, *E. tenella* antigen tc-7a was identified.

CC Sequence 543 BP; 229 A; 135 C; 161 G; 18 T;

Query Match	2.3%	Score 52	DB 1	Length 543
Best Local Similarity	52.3%	Pred. No. 0.00026		
Matches 115	Conservative	0	Mismatches 105	Indels 0
QY 995	accgagacgagcttgcagcagcgcgcgcacgcgtgagcttgcagcgcgcagagagcttcacatgc	1054		
Db 141	AACACACACAGCAACAGCAGCTGCAACAGCAGCAGCTACAGCAACACAGCTGCAGCTGC	200		
QY 1055	agaagcgcgtctgcgtatgtagtccaaacaaatcctccaggaagcagcaggaagatgtgaagcgcgc	1114		
Db 201	AGCACACACTGTCAGCAGCAGCAGCTGTCAGCAGCAGCAGCAGCAGCAACAGCAGCTGCAGCAGC	260		
QY 1115	aaagagaaagaagaattctcccaagaagcagcagaggaataatgtagaagataaccggaagagaga	1174		
Db 261	ACCACCTCAGCAGCAGCAGCGCCAGCAGGAGCAGCAGCAGCAGCAGCAAGAGCGCGAAGACAG	320		



Mon May 22 16:59:56 2000

us-09-502-945-4.rng

Page 15

**QY**    1175    tggacagattgtagaggagaagttaagaagca    1214  
               |                ||| ||| ||| ||| ||| ||| ||| ||| |||  
**Db**    321    AACAGAAACCGAAGGAGCACCAGACGACGACGAGAACA    360

Search completed: May 20, 2000, 00:41:50  
Job time: 12162 sec

**THIS PAGE BLANK (user)**



RESULT 1  
AF039700 2236 bp mRNA PRI 05-MAY-1999  
LOCUS Homo sapiens antigen NY-CO-38 (NY-CO-38) mRNA, complete cds.  
DEFINITION AF039700  
ACCESSION AF039700.1 GI:3170199  
VERSION .  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
AUTHORS Scanlan,M.J., Chen,Y.T., Williamson,B., Gure,A.O., Stockert,E.,  
1 (bases 1 to 2236)  
Gordan,J.D., Tureci,O., Sahlu,U., Pfrenderschu,M., and Old,L.J.  
TITLE Characterization of human colon cancer antigens recognized by  
autologous antibodies  
JOURNAL Int. J. Cancer 76 (5), 652-658 (1998)  
MEDLINE 98272252  
REFERENCE 2 (bases 1 to 2236)  
AUTHORS Scanlan,M.J.  
TITLE Direct Submission  
JOURNAL Submitted (23-DEC-1997) Ludwig Institute for Cancer Research, New  
York Branch at Memorial Sloan-Kettering Cancer Center, 1275 York  
Avenue, New York, NY 10021, USA  
FEATURES  
source  
1..2236  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
gene  
1..2236  
/gene="NY-CO-38"  
CDS  
97..2055  
/gene="NY-CO-38"  
/note="PDZ domain protein"  
/codon\_start=1  
/product="antigen NY-CO-38"  
/protein\_id="AAC18049.1"  
/translation="MDRKVAREPRHKVDFLENDAREKDYLDVLRMYHOTMDVAVLVG  
DKLTVNEPSRLPLFDALRPLPLKHQVEVDLTPPARSKIKRIVRLINLPRSLGISTV  
RGLEFGCELFISHLIKGQADSVGLQVDEYLRINGYSISSTHEEVLINLPRKTYV  
SLKVRHIGLIPVKSPPDEPLTWQYVQFVSESGVSGSLGSPENRNKSKKVFISLVG  
SRGLGSISSGPIQKPGIFISHYKPSGLAEVGLIGDQIVEVNGVDFSNLHKAEVN  
VLKNSRSLTISYAAAGRELFTMDRERLAEARQRELRQBELMOKRLAMESNKIIQEQ  
OEHRORREKKEIAOKAENERYKEMEOIYEERKFKKMEEDMGSKEDLLPKTTTA  
EYHVPRLRKPKYQGVPELEPADLDGGEDEGDFPKRYEEDPYGEMTPPEOIMG  
KQVRLRIKKEGSLDALGSDVSPIGKYVVSAYVREGAERRGGLYKDEIMAINIGK  
IYVDYTLAEADALQKAMNQGDMIDLVAVCPKREYDDELFLILSKRGNQTHALGN  
SELRLPLVNTKPTPSLERGHMTTRHMPMDLNLSPRNLLPLALNQGQIRNSSGHFFE  
GCGGKGGAASRLGEDLKDPSDSHSPLAO"  
BASE COUNT 548 a 588 c 653 g 447 t  
ORIGIN  
Query Match 100.0%; Score 2236; DB 40; Length 2236;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 gtcacatgaaccacgcgtctgcctctgtttgaatgccaatcgcgcgtatcccaatg 300  
|||||  
Db 241 gtcacatgaataaccacgcgtctgcctctgtttgaatgccaatcgcgcgtatcccaatg 300  
QY 301 aagcaccaggttggaataatgatacagctgaacccccgcgcgtccaggaagctgaagagtg 360  
|||||  
Db 301 AAGCACCAGGTGGAATATGATCAGCTGACCCCGCGGCTCCAGGAAGCTGAAGAGAGTG 360  
QY 361 cgtctgagaccgtctgaacccccgaagcctgcctgaatgctggtggtggccggagatt 420  
|||||  
Db 361 CGTCTGAGACCGTCTGACCCCGGAAGGCTCGCCTGAGTGTGCTGCTGGAGATT 420  
QY 421 ggtctgagctcttcacatccacatcaataaagcgtgaagcagacacgcgtgcgc 480  
|||||  
Db 421 GCGCTGAGCGCTTCATCTCCACACCTCATTAAGAGCGGCTGAGGACAGACACGCGGGGCTC 480  
QY 481 caagtaaggagcagaatgcgtcagatcaatgatatcaatcctcctcctgtaccatgag 540  
|||||  
Db 481 CAGGTAGGGGACGAGATCGTCCGGATCAATGATATTCATCTCTCTTACCCATGAG 540  
QY 541 gaagtaatacactcaatctgaacccaagaacatggttccataaagtggaccatcgcgc 600  
|||||  
Db 541 GAGGTATCACTCACTCATTCGAACCAAGAAAACTGTCTCATCAAGTGAACACATCGGC 600  
QY 601 ctgaccccgctgaacagctcctctgaatgaacccctcaactggaatggtgatacagtt 660  
|||||  
Db 601 CTGATCCCGCTGAAAAGCTCTCTGATGAGCCCTCATCTGGCAGATGTGATGATGATTT 660  
QY 661 gtgtcggaaatcgtggggcgtgtcgaagcagcgtggtcctcccttgaaatcgggaaaaa 720  
|||||  
Db 661 GTGTGGAATCTGGGGCGGTGCGAGCAGCCTGGGCTCCCTTGAAATCGGGAACAAG 720  
QY 721 gaaagaagaagttctcaatcagcctgtaagctcccgagggcctgtgcgaagcatctcaag 780  
|||||  
Db 721 GAAAGAAGAGTCTTCATCACCTGTGATGCTCCCGAGGCTTGCTGTCAGCATTTCCAG 780  
QY 781 ggcgccatcagaagcctgtgcatcttatatcagcaatgtgaacatgtgctcctgtcgt 840  
|||||  
Db 781 GGCCCATCCAGAAAGCTGTGCATCTTTATGACGACATGTGAACCTGGTCCGTCTGCT 840  
QY 841 gaagctggatctgaatgaaggaacagatgtcgaagtcaatgctgcgtcctcctaac 900  
|||||  
Db 841 GAGGTGGGATTCGAGATAGGAGCACAGATTGTGAAATGCTCGATCGCTCTCTAAC 900  
QY 901 ctgatacagaagaagctgtgaatgtgtcgtgaataaataagccgcgtccatctccatt 960  
|||||  
Db 901 CTGATCACAAGAGGCTGTAAATGTGCTGAATAAATAGCCGACGCTGACCATCTCATTT 960  
QY 961 gtagctgcaactgctgcggagctgttcatacagacacggagcgtgtgcgaagacgcgg 1020  
|||||  
Db 961 GTAGCTGCACTGCGCGGAGAGCTGTTTCATGACAGACGGGAGCGGCTGCGAAGGCGCG 1020  
QY 1021 cagcgtgaagctgcgcgcgcgaagagcttctcaatgcagaagcgcgtgcgagtgaatc 1080  
|||||  
Db 1021 CAGCGTGAGCTGCAGCGGCGAGAGCTTTCATCATCAAAAGCGGCTGCGAGTGTCAAC 1080  
QY 1081 aagatcctccaggaagcagcagaatgagatgagcgcgcaagaagaaataattgcccaag 1140  
|||||  
Db 1081 AAGATCCTCCAGGAGCAGCAGGAGATGAGCGGCAAGGAAATAATGCGCCGAGAG 1140  
QY 1141 gcagcagaagaaatagagatatccggaagagatatgaaacagatgtaagaggaagaaag 1200  
|||||  
Db 1141 GCAGCAGAGGAAATAGAGATACCGGAAGAGATGGAACAGATTGTAGAGAGGAAGAG 1200  
QY 1201 aagtttaagaagaatggaagaagacgtgggctcaagaagacagctactctgtcctaa 1260  
|||||  
Db 1201 AAGTTTAAGAAGCAATGGGAAGACTGGGCTCAAGAAGACAGCTACTCTTGCTTAA 1260  
QY 1261 accataactgctgaagtaaccacagtaacccttcggaagccaaagtataatgaagagtg 1320  
|||||  
Db 1261 ACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGTATGATCAGGAGGTG 1320



[illegible]

Qy	1448	aggatgctccggctccctcaagcataaagaagagatccttgaacctgagccctggaaagcg	1507
Db	1441	AGGATGTCGGCTCTCTACGATATAGAGAGAGGAGATCCTTAGACTGGCCCTGGAAAGCG	15006
Qy	1508	gtctgagctcccccatttgggaagtgctgcttctctgctgtgatagcggagactgctg	15678
Db	1501	GTGTGGAGCTCCCATATGGAGAGTGGTGTCTTCTCTGTATGAGCGGGAGCTGCTG	15606
Qy	1568	agcgagcatgtgtgcatctgttgaaggggacgagatcatgtgcaatcaacggcaagatltgta	16272
Db	1561	AGCGGCAATGGGCAATTTGTAAAGGGGAGACAGATCATGCAATCAACGGCAAGATTGTGA	16200
Qy	1628	cagactacaccttgctgtaagctgtaagctgtcccttgagaagccttggaaatcagggcggg	16878
Db	1621	CAGACTACACCTTGCTGCTGAGAGCTGTAGGCTGCTCCGACGAAGCCTGTGAAATCAGGGCGGG	16806
Qy	1688	actgagatcgacctgtgtgtgttcgcgctctgcccccgaagagatagatgagatgagctgacct	17472
Db	1681	ACTGATCATGCACTTGTGTGTGGTGGCTGTGCCCCCAAGAGATGATGATGATGAGCTACT	17400
Qy	1748	tctgtctgaagtcacaagggggaacaaatlcacgctttagaagaacagttagctccggc	18078
Db	1741	TCTT-CTGAAGTCCAAAGAGGGGAACCAAAATTGAC-CGTTAGGAACATGAGCTCCGCG	17988
Qy	1808	cccaactctgtgaacacaaagccttcggagcagcctltagaagagccacatgacacacaca	18678
Db	1799	CCCACTCTGTGAACACAAAGCCTCGGATGAGCTTGTAGAGAGGCCACATACACACCA	18588
Qy	1868	gatgtgactcttggagacctgaactctcaacccagagatctaaatccctcttggccctga	19272
Db	1859	GATGGCAATCCTTGGGACCTTAATCTATCACCCAGGATCTCAAACTCCCTTTGGCCCTGA	19188
Qy	1928	accgagggccagataaagaacagcttcggccacttlttlaaaggccaaltgttggagaaag	19878
Db	1919	ACCAAGGCCAGATATAGGAACAGCTGGGGCCACTCTTGTGAAGGCCAACGTGGAGAAAGG	19788
Qy	1988	gagcagccagcgcttttggggagaagatctcaaggaatccagactctcaatccttctctctg	20472
Db	1979	GAGCAGGCACCAATTTGGGAGGAAGATCTCAAGGATCAACACTCTATTCCTTCTCTGAGG	20388
Qy	2048	ccgagtgaaattgtctctccagcttlttggggagactcttctcttgaacctataagacc	21078
Db	2039	CCCACTGAATTTGGTCTCTCCACAGCTCTGGGGGACATCTTCTTGTGAACCTATATAGACC	20988
Qy	2108	ccactgtagctctctctctcatalccctctctctctgcccctcgacttaatgtgtccag	21678
Db	2099	CCACTGGAGTCTCTCTCTCCATCCCTCTCCTCTGCTGCTATATATCTCTCCAGG	21588
Qy	2168	attgtcaactccaaccttaact	2188
Db	2159	ATTGTCACTCCAAACTTACT	2179
RESULT	3		
LOCUS	AF039699	2162 bp	mRNA PRI 05-MAY-1999
DEFINITION	Homo sapiens antlygen NY-CO-37 (NY-CO-38) mRNA, complete cds.		
ACCESSION	AF039699		
VERSION	AF039699.1	GI:3170197	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
AUTHORS	1 (bases 1 to 2162) Scanlan,M.J., Chen,Y.T., Williamson,B., Gure,A.O., Stockert,E., Jordan,J.D., Tureci,O., Sahin,U., Pfleundschuh,M. and Old,L.J.		
TITLE	Characterization of human colon cancer antigens recognized by autologous antibodies		
JOURNAL	Int. J. Cancer 76 (5), 652-658 (1998)		
MEDLINE	98272252		
REFERENCE	2 (bases 1 to 2162)		
AUTHORS	Scanlan,M.J.		



QY	1681	gaccggagactgacatgcacctgctgtgtccgctgtcccccacaaaggatatagcagatgag	1740
Db	1607	GGCGGGGAACTGGATGACACTTGTGTGTTCCGCTCTGCCCCCAAAGAGATGTGACATGAG	1666
QY	1741	ctgacctctctgtcgtgaagtcctcaaaaggggaaacccaatctacgcgtttaggaacagtgag	1800
Db	1667	CTGACCTCTTCTGTGAATCTCCAAAAGGGGAAACCAAAATTCACCCGTTAGAAACAGTGAG	1726
QY	1801	ctccggcccacactcgtttaacacaaagccttcgagacccgctttagagagagccacatgaca	1860
Db	1727	CTCCGGCCCCACCTCTGTAACACAAAGCTTGAGACACGCTTGAAGAGGCCACATGATA	1786
QY	1861	cacaccagatgycalcctctggagacctgaatctatcacccagaaatctcaaatccctctg	1920
Db	1787	CACACCAATGGCATCTTGGAGCCTGAATCTATCACCCAGGAATCTCAACCTCCCTTGG	1846
QY	1921	gccctgaaccagggccagataaagaacagctcgggccactttttgaagggccaatgtgga	1980
Db	1847	GCCCTGAACCAAGGGCCAGATTAAGGAAACGTCGGGGCACTTTTAAAGGCCCAATGTGGA	1906
QY	1981	ggaagggagagcagccagccgcttgggagaagatctcaagaatccagacactcatccctt	2040
Db	1907	GGAAGGGAGACGACGACGCCCTTTTGGGAAAGATCTCAAGATTCACACTCTAATTCCTT	1966
QY	2041	cctctgcccagtgaaattggtctctctccacgcttggggagactcctctctgaacctaa	2100
Db	1967	CCTCTGCGCCAGTAAATTTGGCTCTCTCCAGCTTTGGGGAGACTCTTCTTGACCCCTAA	2026
QY	2101	taagaccccatggaagctctctctctccatccctctctctctgacctctgctctctaatgc	2160
Db	2027	TAAAGCCCCACTGGAAGTCTCTCTCTCCACATCCCTCTCTGCCCCCTGCTCTAATTGC	2086
QY	2161	tgccagatgtcactccaaacctactcctgagctcatataataaaataacagattat	2220
Db	2087	TGCCAGGATGTCTCATCTCCAAACCTTACTGTGAGCTCATTAATAATAACAGATTAT	2146
QY	2221	ttccaagcttaaaaaa 2236	
Db	2147	TTCCAGCTTAAAAAAA 2162	
RESULT 4			
LOCUS	AB006955		
DEFINITION	AB006955 2283 bp mRNA	PRI	02-DEC-1999
ACCESSION	AB006955		
VERSION	AB006955.1 GI:5152287		
KEYWORDS	AIE-75; aie-75.		
SOURCE	Homo sapiens small intestine cDNA to mRNA.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.		
	1 (sites)		
	Kobayashi, I., Imanura, K., Kubota, M., Ishikawa, S., Yamada, M.,		
	Tonoki, H., Okano, M., Storch, W. F., Moriwuchi, T., Sakiyama, Y. and		
	Kobayashi, K.		
TITLE	Identification of an autoimmune enteropathy-related 75-kilodalton		
	antigen		
JOURNAL	Gastroenterology 117 (4), 823-830 (1999)		
MEDLINE	99431904		
REFERENCE	2 (bases 1 to 2283)		
AUTHORS	Kobayashi, I		
TITLE	Direct Submission		
JOURNAL	Submitted (31-AUG-1997) to the DDBJ/EMBL/GenBank databases. Ichiro		
	Kobayashi, Hokkaido University School of Medicine, Department of		
	Pediatrics, North-15, West-7, Kita-ku, Sapporo, Hokkaido 060-8638,		
	Japan (E-mail: ichikobadamed.hokudai.ac.jp, Tel:011-716-1161,		
	Fax:011-736-9267)		
FEATURES			
SOURCE	Location/Qualifiers		
	1..2283		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		

gene	/tissue_type="small intestine"
246. .1811	
/gene="aie-75"	
246. .1811	
/gene="aie-75"	
/codon_start=1	
/product="AIE-75"	
/protein_id="BAA81739.1"	
/db_xref="GI:5152288"	
/translation="MYHQMDEVAVLVGDLKLVINPRLPFLDAIRPLPLKHQVEYI QLTPRSKRFLKVELDRLHPBELGSLVSQGLEGGGLTSLIKGQADSVGLQVGD LYRINGYSISSTCHBEVNLINPTKTVSLKXNIGLIPKSSPDEPLMVOYDPS SGGVGSLSPGPNRENKKEKVFSLVSGSKGLSCSLISGPIQKGFISHVPGSLSA VLELGIDQIVEYNGVFNLDHKEAVNYVILSRSLLTSLVAAGSELMTDEIRTELE RROELROELMOKRLAMESNKLIOEODEMERORRKEIAOKRAEENYERMEQIV PEERKTKOMEDMGSKEOLLPKTTAYVHPVLPKPYDOCVPEELPADIDSGTE EGGEDPFRKYBERGDPPIYSNFTPEQJMGKDVRLIRIKKESLDLALGVDSEFIGKVV SAVYERGAERHGIIVKGDDEIWAINGKIVLYDTTLAEAPRALOKAWNQGMDILVAA CPKREYDEDLRTFF"	



Db 847 AGAAGCCTGGCATCTTATATCAGCCATGTGAACCTGGCTCCCTGTCTGTAGGTGGAT 906  
 QY 851 tggagataaggagaccagattgtcgaagtcaatggcgtcgactctcctaaccgtgatca 910  
 Db 907 TGGAGATAGGGGAGACCAATGTGCGAAGTCAATGCGTGCATCTCTTAACCTGGATACA 966  
 QY 911 aggaagctgtaaatgtctgtaaaaaatagccgcagcctgacacatcgcattgttagctcag 970  
 Db 967 AGGAGGCTGTAAATGTCTTGAAGATGAGCCGACCTGACCATCTTCCATTGTACTCTGAG 1026  
 QY 971 ctggccggagagctgttcaatcacagaccggagcgctgacagagcgccggcagctgac 1030  
 Db 1027 CTGGCCGGGAGCTGTTCATGACAGACCGGGCGGTGACAGAGCGCGGAGCGGTGAGC 1086  
 QY 1031 tgcagcgccagagagctctcatcagaagcgctgacatgagatccacaagaatcctcc 1090  
 Db 1087 TGGAGCGCGGAGGAGCTTCTATGCGAAGCGCGCTGGCGATGAGATCCAAACAGATCTCC 1146  
 QY 1091 aggaagcagagagatgtagcgagaaagaaagaaatgtgcccagaagcgagagg 1150  
 Db 1147 AGGAGCGAGAGATGAGACCGCAAGAGAAAGAAATTTGCCAGAGGACAGAGAGG 1206  
 QY 1151 aaaaatgagataccggaagagagatggaacagattgtagaagagaaagagaaagtttaaga 1210  
 Db 1207 AAATATGAGATATACCGGAGAGATGGAACAGATTGTAGAGAGAGAGAGAGATTTAAGA 1266  
 QY 1211 agcaatggaagaagacttgggtctcaagaaacagctactcttgcttaaaacatactg 1270  
 Db 1267 AGCAATGTGGAAAGAGTGGGGGCTCAAAAGAACAGCTACCTTCCCTAAAACCATCACTG 1326  
 QY 1271 ctgaagtgacacccagatcccccttcgcaagcagaagtatgtaagagaggttgaagctagc 1330  
 Db 1327 CTGAGGTACACCCGGTACCCCTTGGCAAGCCAAAGATGATGACAGGGAGTGAACCTGAGC 1386  
 QY 1331 tcgagcccgacagatgacatgtagatgtagcagcagagagacagagagagattccgga 1390  
 Db 1387 TCGAGCCCGGAGATGACCTGGATGAGAGCAAGGAGAGAGAGAGAGAGAGATTTCGGGA 1446  
 QY 1391 aataatggaagagctttagccctcactatgttcaacccagagacagatcatggtggaag 1450  
 Db 1447 AAATATGAGAAAGGCTTGTGACCCCTACCTATGTTACACCCAGACAGATCATGCGGAGG 1506  
 QY 1451 atgtccggctcctaagcataaagaagaagagatccttaagacctggccctggaagcggtg 1510  
 Db 1507 ATGTCCGGCTCTATCGCATGAAGAGAGGAGATCTTAGACCTGGCCCTGGAAGCGGCTG 1566  
 QY 1511 tggactcccccatggaaggtgtcgtcttcctgtgtatgtagcgaggagctctgac 1570  
 Db 1567 TGGACTCCCCCATTTGGGAAGGTGCTTCTTCTGTGTATGAGCGGGGAGCTGCTGAGC 1626  
 QY 1571 ggcattgtgcatgtgtgaaaggagagacagatcatgagcaatcaacgagcaagattgtgacag 1630  
 Db 1627 GGCATGTGTGCATTTGTAAGGGGAGAGATCATGCAATCAACGGCAAGATTGTGACAG 1686  
 QY 1631 actaacacctgctgagagctgagcgctgacgctgcaagagcgctggaatcagggcgaggact 1690  
 Db 1687 ACATACACCTGCTGAGAGGCTGAGGCTGCTCCCTGCGAAGAGGCTGGAATCAGGGCGGACT 1746  
 QY 1691 ggaatgcacctgtgtgttcgcctgtgcccccaagagagatgtagagatgtagcctctc 1750  
 Db 1747 GGATCGACCTTGTGTGCTGCGCTGCGCCCAAGAGGTATGACGATGAGCTGACTTCT 1806  
 QY 1751 tgcctgaagctcaaaaggagaaaccaatltcaagcgtttagaagaacagtgagctccggcccc 1810  
 Db 1807 T-CTGAAGTCCAAAGGGGAAACCAATTCAC-CGTTAGGAAACAGTAGAGTCCGGCCCC 1864  
 QY 1811 acctcgtgaaacaaagcctcgagcccgcttggagagagcgcaataaacaacaacagat 1870  
 Db 1865 ACCTCGTGAACACAAACCTCGGTGATCAGCCTTGGAGAGGGCCACATCAACACACGAT 1924  
 QY 1871 ggcattccttggagactaatcatcacccaggaatcacaactccttggcccttgacc 1930  
 Db 1925 GGCATCTTGGGACCTTAATCTATACCCAGGAATCTCAAACTCCCTTTGGCCCTGAAC 1984

QY 1931 agggccagataaagaacagctcgggccactttttgaaaggccaatgtgagagaaaggag 1990  
 Db 1985 AGGGCCAGATAAAGGAACAGCTCGGGCCACTTCTTGAAGGCGCAACCTGAGGAAGGAG 2044  
 QY 1991 cagccagccgttggagaaagatcacaagatcagaactcactcctccttccttcctgc 2050  
 Db 2045 CAGCCACCATTTGGGAGGAAGATTCAGAGATCCAGACTCTCATCTCTTCTCTTGCCCC 2104  
 QY 2051 agtgaattgtctctccagcttggggagctcctcttgaacccttaataagaccaca 2110  
 Db 2105 AGTGAATTTGGTCTCTCCAGCTGTGGGGA-TCCTTCTCTTGAACCTTAATGAACCCCA 2163  
 QY 2111 ctggaagctctctctccatccctcctcctcctcctcctcctcctcctcctcctcct 2170  
 Db 2164 CTGAGATCTCTCTCTCCACTCCCTCTCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 2223  
 QY 2171 gtacacccaacctactctgagctcatataaataaacaagattatcttcagctta 2230  
 Db 2224 GTCACTCCAAACCTTACT-----GCACTCATGGAATAAATTAATTAACCACTAA 2277  
 QY 2231 aaaaaa 2236  
 Db 2278 AAAAAA 2283  
 RESULT 5  
 AC005137/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 HTG.  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 161190)  
 Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,  
 Buettner,J., Bumeister,R., Card,P., desaliboat,C., Dunn,J.,  
 English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,  
 Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N.,  
 Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,  
 Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.  
 HTG Submission  
 Direct Submission  
 Submitted (19-JUN-1998) Genome Science & Technology Center,  
 University of Texas Southwestern Medical Center, 5323 Harry Hines  
 Blvd, Dallas, TX 75235-8591, USA  
 3 (bases 1 to 161190)  
 Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,  
 Buettner,J., Butler,C., Card,P., desaliboat,F., Dunn,J.,  
 English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,  
 Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N.,  
 Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,  
 Schageman,J., Schultz,R.A., Stimson,S., Waller,K. and Ward,T.  
 Direct Submission  
 Submitted (30-SEP-1998) Genome Science & Technology Center,  
 University of Texas Southwestern Medical Center, 5323 Harry Hines  
 Blvd, Dallas, TX 75235-8591, USA  
 COMMENT  
 Further information regarding the map of this region or  
 annotation of 6-106f23 can be found at  
 http://gsc.tsc.tamuedu/chromosome2.htm.  
 CHROMOSOMAL LOCUS: This PAC clone comes from the Usher syndrome

region (USHC) mapped between STS markers D1S1310 and 11S1A14.  
MARKER CONFIRMATION: STS/ESTs sequence confirmed; D1S1032 and D1S4138.  
MAPPED CLONE OVERLAP: pDU239b22 and 6-109b6.  
IMPORTANT: This submission contains the entire insert of clone 6-106f23. 6-106f23 comes from a PAC library constructed at the Roswell Park Cancer Institute by the Pieter de Jong group. This clone has been finished according to strict quality criteria and attempts have been made to resolve all base calling problems such as compressions and repetitive elements. The expected Phred/Phrap calculated errors/10kb is 0.40. In addition, this sequence has been finished such that 99.9% of consensus base calls consist of either double-stranded coverage or 2 types of labeling chemistry on one strand.

```
FEATURES
Source          Location/Qualifiers
repeat_region   1..161190
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /rpt_family="MLT282"
repeat_region   574..955
                /rpt_family="MIR"
repeat_region   1445..1635
                /rpt_family="MIR"
repeat_region   3125..3302
                /rpt_family="MIR"
repeat_region   11175..11329
                /rpt_family="MIR"
repeat_region   14552..14716
                /rpt_family="MIR"
repeat_region   complement(18589)..18627
                /rpt_family="MIR"
repeat_region   23165..23232
                /rpt_family="MIR"
repeat_region   complement(25103)..25361
                /rpt_family="Alu"
repeat_region   29622..29909
                /rpt_family="Alu"
repeat_region   complement(32772)..32989
                /rpt_family="Alu"
repeat_region   33175..33284
                /rpt_family="MIR"
repeat_region   complement(34762)..34852
                /rpt_family="MIR"
repeat_region   34864..34996
                /rpt_family="MIR"
repeat_region   complement(35719)..35843
                /rpt_family="Alu"
repeat_region   complement(36294)..36467
                /rpt_family="MIR"
repeat_region   36817..37084
                /rpt_family="Alu"
repeat_region   37953..38186
                /rpt_family="MIR"
repeat_region   39000..39177
                /rpt_family="MIR"
repeat_region   complement(39898)..39968
                /rpt_family="MIR"
repeat_region   complement(41579)..41780
                /rpt_family="MIR"
repeat_region   complement(43317)..43597
                /rpt_family="Alu"
repeat_region   45609..45899
                /rpt_family="Alu"
repeat_region   complement(46819)..46884
                /rpt_family="MIR"
repeat_region   complement(48840)..48895
                /rpt_family="MIR"
repeat_region   50385..50543
                /rpt_family="MIR"
repeat_region   complement(50842)..51126
                /rpt_family="Alu"
repeat_region   complement(53140)..53180
                /rpt_family="MIR"
repeat_region   complement(55758)..56060
```

```
repeat_region   /rpt_family="Alu"
                complement(59724)..60021
repeat_region   /rpt_family="Alu"
                66550..66645
repeat_region   /rpt_family="MIR"
                73180..73260
repeat_region   /rpt_family="Alu"
                complement(74694)..74990
repeat_region   /rpt_family="Alu"
                complement(76214)..77037
repeat_region   /rpt_family="MER42"
                complement(77130)..77414
repeat_region   /rpt_family="MER2"
                complement(77532)..77641
repeat_region   /rpt_family="MER42"
                complement(78518)..78630
repeat_region   /rpt_family="Alu"
                complement(78918)..79211
repeat_region   /rpt_family="Alu"
                85547..85650
repeat_region   /rpt_family="MIR"
                88495..88613
repeat_region   /rpt_family="MIR"
                complement(93509)..93900
repeat_region   /rpt_family="L1"
                complement(94148)..94417
repeat_region   /rpt_family="Alu"
                complement(95536)..95822
repeat_region   /rpt_family="Alu"
                complement(96013)..96567
repeat_region   /rpt_family="L1"
                complement(106391)..106583
repeat_region   /rpt_family="MLT1"
                complement(114951)..115136
repeat_region   /rpt_family="MIR"
                118500..118580
repeat_region   /rpt_family="MIR"
                complement(126403)..126593
repeat_region   /rpt_family="L1"
                126665..126942
repeat_region   /rpt_family="Alu"
                complement(127152)..127352
repeat_region   /rpt_family="Alu"
                complement(128074)..128639
repeat_region   /rpt_family="L1"
                complement(128923)..129156
repeat_region   /rpt_family="Alu"
                130009..130278
repeat_region   /rpt_family="Alu"
                complement(134144)..139342
repeat_region   /rpt_family="L1"
                complement(137205)..137469
repeat_region   /rpt_family="Alu"
                complement(140667)..140762
repeat_region   /rpt_family="MIR"
                complement(142407)..142521
repeat_region   /rpt_family="MER42"
                147830..148006
repeat_region   /rpt_family="MIR"
                149821..150066
repeat_region   /rpt_family="Alu"
                complement(151137)..151327
repeat_region   /rpt_family="MIR"
                160309..160771
repeat_region   /rpt_family="MLT1"
                37674 a 40113 c 40195 g 43208 t

BASE COUNT 37674 a 40113 c 40195 g 43208 t
ORIGIN
```

Query Match 19.6% Score 437.4 DB 11 Length 161190  
Best Local Similarity 95.7% Pred. No. 3.7e-90  
Matches 472; Conservative 0; Mismatches 16; Indels 5; Gaps 2;

OY	1740	gcacacctcttgcctgaagtcacaaagggaaccaaattcaagcctttaggaacactga	1799
Db	5253	GCAGGACCTTTCTTGAGATCCAAAAGGGAAACCAATTCAC--CGTTAGAAAACAGTGA	5195
OY	1800	gtctcgggcccacactcgtgtaacacaaagccttcggacccgctctagagagccacatgac	1859
Db	5194	GCTCGGGCCCACTCGTGAACACAAACCTTCGGATCAGCTTGAAGAGGCCACTAC	5135
OY	1860	acacacagatgtgcatctcttggagacctgaatctatcaaccagaatctcaactccctt	1919
Db	5134	ACACACCCGATGGCATCTCTGGGACCTGAATCTATACACCAGAACTCAAACTCCCTTT	5075
OY	1920	ggccctgaaccaaggcccgatagaagaagctcgggcacttttttaagggcgaattgg	1979
Db	5074	GGCCCTGAACCAAGGGCCCGATAAAGAAACGCTCGGGCCACTCTTGAAGGCCAACCTGG	5015
OY	1980	aggaaggaggacagccagccgcttgggaagaagctcaagatccgaacttaactcctt	2039
Db	5014	AGGAAAGGAGACAGCACCACTTTGGGAGAAAGATCTCAAGGATCCAGACTCTCAATTCCTT	4955
OY	2040	tcctctggcccagatgaatttgcctctctccacgcttggggagaccccttccttgaacccta	2099
Db	4954	TCTCTGGCCCAAGTAATTTGGTCTCTCTCCAGCTCTGGGGGACCTCTTCTTGAACCTTA	4895
OY	2100	ataagaccccacitgagctctctctctctcaatccctctctctctgcccctgctctaattg	2159
Db	4894	ATAAGACCCCACTGGAGTCTCTCTCTCCATTCCTCTCTCGCCCTCTCTTAATTTG	4835
OY	2160	ctgcgcaggaattgtcactccaacttactctgtagctcaatataaataaacaagatttat	2219
Db	4834	CTGCAGGATTTGTCACTCCAAACCTTACTGTGAGCTCATTAATAAAT---AGATTTAT	4779
OY	2220	tttcagactaa 2232	
Db	4778	TTTCCAGCTTATA 4766	

RESULT	6
LOCUS	G55504/c
DEFINITION	G55504 463 bp DNA STS 26-OCT-1999
ACCESSION	SHGK-100852 Human Homo sapiens STS genomic, sequence tagged site
VERSION	G55504 G55504
KEYWORDS	G55504.1 GI:6120823
SOURCE	STS.
ORGANISM	human. Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	1 (bases 1 to 463) Myers,R.M.
TITLE	Human STSs (1999)
JOURNAL	Human STSs (1999)
COMMENT	Unpublished (1999)

Contact: Richard M. Myers  
Stanford Human Genome Center (SHGC)  
Stanford University School of Medicine  
Department of Genetics, M-344, Stanford, CA 94305, USA  
Tel: 4157259687  
Fax: 4157259689  
Email: myers@shgc.stanford.edu  
Primer A: TCCAAAGTCAGCAATTCACATCT  
Primer B: TGTAAGAAGCAACGATCATCG  
STS size: 313  
PCR Profile:  
Initial Incubation: 95 degrees C for 10 minutes  
Denaturation: 94 degrees C for 30 seconds  
Annealing: 60 degrees C for 30 seconds  
Polymerization: 72 degrees C for 23 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9700  
Protocol:  
Template: 25 ng  
Primer: each 1 uM

dNTPs:	each 200 uM
AmpliTaq Gold Polymerase:	0.07 units/ul
Total Vol:	5 ul
Buffer:	
MgCl2:	2.5 mM
KCl:	50 mM
Tris-HCl:	10 mM
pH:	8.3
BAC ends sequenced at TIGR from the RPCII BAC library. Designed and developed at the Stanford Human Genome Center.	
FEATURES	
source	Location/Qualifiers
	1..463
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/map_11"
	/clone_11b="Human"
STS	80..392
primer_bind	80..102
primer_bind	complement(370..392)
BASE COUNT	96 a 116 c 128 g 123 t
ORIGIN	

Query Match	4.7%	Score 105	DB 13	Length 463
Best Local Similarity	82.8%	Pred. Nc. 6e-14	Mismatches 25	Indels 0
Matches 120	Conservative	0	Gaps 0	
QY	1576	gctggcattgtttaaaggagacagatcatgycgaatcaacggaagtgtgacagactac	1635	
DB	400	ggtggcattgtttaaaggagacagatcatgycgaatcaacggaagtgtgacagactac	341	
QY	1636	accctgctgaagcctgaagcctgccttcgaagaagccttgtaacgaagcgaggactatgc	1695	
DB	340	accctgctgaagcctgaagcctgccttcgaagaagccttgtaacgaagcgaggactatgc	281	
QY	1696	gacctgtgtgtgtgcgcgtctgcacccc	1720	
DB	280	agggcccttcccttcccttccctc	256	

	RESULT	7				
	LOCUS	I66494	7218 bp	DNA	PAT	28-DEC-1997
	DEFINITION	Sequence	14 from patent US 5670367.			
	ACCESSION	I66494				
	VERSION	I66494.1	GI: 2724471			
	KEYWORDS	.				
	SOURCE	Unknown.				
	ORGANISM	Unknown.				
	REFERENCE	Unclassified.				
	AUTHORS	1 (bases 1 to 7218)				
	TITLE	Donner,F., Scheiflinger,F. and Falkner,F.Gunter. Recombinant fowlpox virus Patent: US 5670367-A 14 23-SEP-1997;				
JOURNAL	Location/Qualifiers	1..7218				
FEATURES	source	/organism="unknown"				
BASE COUNT	1944 a	1491 c	1486 g	1929 t	368 others	
ORIGIN						

```

Query Match          3 6%, Score 81.4: DB 5, length 7218:
Best Local Similarity 4.7%, Pred. No.1.9e-08;
Matches 19; Conservative 246; Mismatches 142; Indels 0; Gaps 0

QY 1032 gcaagcagcagagctctcatcagaagcagctgcgcatggatccaaagaatctcca 1091
      | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1438 GTCACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1379

QY 1092 ggaagcagcagagatggaagcagcaagaagaataattgccagaagcagcagaagga 1151

```

[illegible]

TITLE	COMMENT
Direct Submission	
Submitted (04-JUN-1999) Whitehead Institute for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
On Jun 4, 1999 this sequence version replaced gi:4966400.	
All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)	
http://ftp.genome.washington.edu/RM/RepeatMasker.html	
Only 180385 base pairs from the middle of this clone are being submitted. The remainder overlaps either accession number AC005152 (WICR project L335) or accession number AC005243 (WICR project L343).	
FEATURES	
Source	Location/Qualifiers
repeat_region	1..180385
repeat_region	/organism="Homo sapiens"
repeat_region	/db_xref="taxon:9606"
repeat_region	/clone="B4_E_24"
repeat_region	/clone_lib="Alan Buckler -- per comm"
repeat_region	/map="17"
repeat_region	/chromosome="17"
repeat_region	446..753
repeat_region	/rpt_family="AluSc"
repeat_region	2375..2419
repeat_region	/rpt_family="(TTTG)n"
repeat_region	complement(2985..3103)
repeat_region	/rpt_family="L2"
repeat_region	complement(3146..3352)
repeat_region	/rpt_family="L2"
repeat_region	complement(3531..3639)
repeat_region	/rpt_family="L2"
repeat_region	complement(3796..4394)
repeat_region	/rpt_family="L2"
repeat_region	5531..5631
repeat_region	/rpt_family="MER81"
repeat_region	5648..5674
repeat_region	/rpt_family="(CAAAA)n"
repeat_region	5812..6020
repeat_region	/rpt_family="L1MC4"
repeat_region	6180..6224
repeat_region	/rpt_family="AT-rich"
repeat_region	6303..6611
repeat_region	/rpt_family="AluSp"
repeat_region	7378..7618
repeat_region	/rpt_family="MIR"
repeat_region	9593..9623
repeat_region	/rpt_family="(GGGA)n"
repeat_region	complement(10666..10923)
repeat_region	/rpt_family="AluSq"
repeat_region	10938..11088
repeat_region	/rpt_family="L2"
repeat_region	11752..11914
repeat_region	/rpt_family="MIR"
repeat_region	complement(11915..12219)
repeat_region	/rpt_family="AluX"
repeat_region	12220..12347
repeat_region	/rpt_family="L2"
repeat_region	13110..13182
repeat_region	/rpt_family="GA-rich"
repeat_region	13526..13564
repeat_region	/rpt_family="(TCC)n"
repeat_region	13575..13660
repeat_region	/rpt_family="(TA)n"
repeat_region	complement(14380..14658)
repeat_region	/rpt_family="AluIo"
repeat_region	14742..14830
repeat_region	/rpt_family="(TA)n"
repeat_region	14860..14967
repeat_region	/rpt_family="(TA)n"
repeat_region	complement(14977..15122)
repeat_region	/rpt_family="L2"
repeat_region	complement(15292..15385)
repeat_region	/rpt_family="L2"
repeat_region	16672..16781

```

repeat_region      /rpt_family="L2"
16802. .16846
/rpt_family="(TGAA)n"
repeat_region      17225. .17245
/rpt_family="(A)n"
repeat_region      18411. .18720
/rpt_family="AluB"
repeat_region      complement(20715. .20788)
/rpt_family="MLTJ2"
repeat_region      complement(20822. .20916)
/rpt_family="MLTJ2"
repeat_region      20956. .21137
/rpt_family="MIR"
repeat_region      21830. .21872
/rpt_family="AT_rich"
repeat_region      21957. .22257
/rpt_family="AlusX"
repeat_region      complement(22438. .22538)
/rpt_family="MLTJ"
repeat_region      complement(23375. .23667)
/rpt_family="AluY"
repeat_region      24553. .24858
/rpt_family="AluY"
repeat_region      complement(25021. .25234)
/rpt_family="MIR"
repeat_region      26957. .27261
/rpt_family="AluY"
repeat_region      complement(27524. .27669)
/rpt_family="MLT1B"
repeat_region      complement(27670. .27964)
/rpt_family="AlusP"
repeat_region      complement(27965. .28210)
/rpt_family="MLT1B"
repeat_region      complement(28228. .28522)
/rpt_family="AlusG"
repeat_region      complement(28592. .28700)
/rpt_family="MIR"
repeat_region      complement(29353. .29537)
/rpt_family="MIR"
repeat_region      29802. .30095
/rpt_family="AlusG"
repeat_region      complement(30099. .30219)
/rpt_family="MIR"
repeat_region      30227. .30267
/rpt_family="(TG)n"
repeat_region      30276. .30321
/rpt_family="(CA)n"
repeat_region      complement(30322. .30611)
/rpt_family="AluJo"
repeat_region      complement(30956. .31151)
/rpt_family="MER3"
repeat_region      complement(31533. .31633)
/rpt_family="L2"
repeat_region      31810. .31858
/rpt_family="L2"
repeat_region      32028. .32107
/rpt_family="L1PA4"
repeat_region      complement(32919. .33182)
/rpt_family="MIR"
repeat_region      34169. .34314
/rpt_family="CT-rich"
repeat_region      34774. .35007
/rpt_family="MIR"
repeat_region      35589. .35621
/rpt_family="AT_rich"
repeat_region      35763. .36142
/rpt_family="MER41A"
repeat_region      36143. .36452
/rpt_family="AlusG"
repeat_region      36453. .36636
/rpt_family="MER41A"
repeat_region      38011. .38071
/rpt_family="CT-rich"

```

```

repeat_region      38083. .38262
/rpt_family="(TTC)n"
repeat_region      38235. .38319
/rpt_family="(TTC)n"
repeat_region      38320. .38499
/rpt_family="(TTC)n"

Query Match      3.1%; Score 69.6; DB 40; Length 180385;
Best Local Similarity 53.7%; Pred. No. 1.2e-05;
Matches 144; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

Qy 978 ggaagctgtcattgacagccgagcgctgtgcagagcgcgagcgctgtgcagcg 1037
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38451 GGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38392

Qy 1038 gcaagagcttcattcagcaagcgctgtgcagagcgctgtgcagagcgca 1097
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38391 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38332

Qy 1098 gcaagagatgtgagcgcgcaagagagaagaattgtccagagagcgagagaaatga 1157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38331 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38272

Qy 1158 gagtaccggaagagagatgtgacagattgtagagagagagagagagagagagcaat 1217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38271 GGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38212

Qy 1218 ggaagagagctgtgtgctcaagagagag 1245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38211 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38184

RESULT 9
AC016063          151428 bp      DNA      HTG      14-FEB-2000
LOCUS             Homo sapiens clone RP11-1706, WORKING DRAFT SEQUENCE, 12 unordered
DEFINITION        pieces.
ACCESSION          AC016063
VERSION            AC016063.3 GI:6970444
KEYWORDS           HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE             human.
ORGANISM           Homo sapiens
                   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                   Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE          1 (bases 1 to 151428)
AUTHORS            Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE             Homo sapiens, clone RP11-1706
JOURNAL            Unpublished
REFERENCE          2 (bases 1 to 151428)
AUTHORS            Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
                   Baldwin,J., Barina,N., Beckerly,R., Boguslavsky,L., Bouknight,B.,
                   Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
                   Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,D., Doyle,M.,
                   Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
                   Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,
                   Howland,J.C., Johnson,R., Jones,C., Kann,L., Karszas,A., Klein,J.,
                   Lehoczy,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,
                   McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
                   Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
                   Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
                   Strange-Rhomon,N., Stojanovic,N., Subramanian,A., Talamas,J.,
                   Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
                   Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
JOURNAL            Direct Submission
COMMENT            Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome
                   Research, 320 Charles Street, Cambridge, MA 02141, USA
                   On Feb 14, 2000 this sequence version replaced gi:6721365.
                   All repeats were identified using RepeatMasker:
                   http://ftp.genome.washington.edu/RM/RepeatMasker.html
                   ----- Genome Center
                   Center: Whitehead Institute/ MIT Center for Genome Research
                   Center code: WIBR

```



*	12752	15275:	contig of 2524 bp in length
*			gap of unknown length
*	15276	20080:	contig of 4805 bp in length
*			gap of unknown length
*	20081	24494:	contig of 4414 bp in length
*			gap of unknown length
*	24495	31437:	contig of 6943 bp in length
*			gap of unknown length
*	31438	37373:	contig of 5936 bp in length
*			gap of unknown length
*	37374	44749:	contig of 7376 bp in length
*			gap of unknown length
*	44750	50716:	contig of 5967 bp in length
*			gap of unknown length
*	50717	56335:	contig of 5619 bp in length
*			gap of unknown length
*	56336	66076:	contig of 9741 bp in length
*			gap of unknown length
*	66077	76677:	contig of 10601 bp in length
*			gap of unknown length
*	76678	94859:	contig of 18182 bp in length
*			gap of unknown length
*	94860	118422:	contig of 23563 bp in length
*			gap of unknown length
*	118423	142659:	contig of 24237 bp in length
*			gap of unknown length
*	142660	166171:	contig of 23512 bp in length

FEATURES	Location/Qualifiers
source	1. .166171

	/clone_11b="RPC1-11 Human Male BAC"									
BASE COUNT	49124	a	37104	c	36032	g	42814	t	1097	others
ORIGIN										

Query Match	3.18	Score 68.8	DB 42	Length 166171
Best Local Similarity	52.98	Pred. No. 1.8e-05		
Matches 148; Conservative	0	Mismatches 132	Indels 0	Gaps 0

DQY 976 cggagactgtcatcacagaccggaagcgcgtgcagaaggcgccgcacgttgaqctgcag 1035  
||||| | | ||||| | | ||||| | | ||||| | | |||||  
Db 130472 CCGGAGACGAGGGGGCAGTGC GGGAAGCACAGAGGGCCAGCTGTCCGAGACAGGAGGGCCAG 1305

QY 1036 cggcagagagcttctcatgcaagcgcgcgtgcgatgagttccacaagatcctcagagag 1095  
 ||||| ||||| | | | | | |||||  
 Db 130532 GTCTCGGAGACGAGAGGGCGATGGGGGAGACGAGAGGGCGAGATGGGGGAGCAGAGGAG 130591

Dy 1096 caccaggaatgtagcggcaagaataaatctccagaaggcacgacgaanaat 1155  
||| ||| | | | | | | | | | | |  
Db 130592 CAAATGGGAGCAGGNGACAGAAGCAAGCAGAGGAGACAGATGGGGACCAGGA 130671

QY 1156 gaagataccggaaggagatgtgaacagattgtagaagaggaagaagttaagaaccaa 1215  
||| ||||||| ||| ||||| ||| ||||| ||| ||| ||| |||||  
Db 130652 GAGCAGATTCGGAACAGACAGAGAGATGGGGGAGACAGAGAGAGCGAGGTGCAGAAGCAG 1307

Oy 1216 tgggaagaagactggygctcaaggaacagctactcttgc 1255  
 ||| | | | | | | |  
 Db 130712 GAGGAGCAGGCTCAAAAGCAGGAGGAGCAGATGGCGCAAG 130751  
 ||| | | | | | | |

## RESULT 11

LOCUS	3489 bp	DNA	24-OCT-1996
DEFINITION	Kaposi's sarcoma-associated herpes-like virus ORF73 homolog gene, complete cds		

```

ACCESSION 052004
VERSION   052064.1
KEYWORDS  GI:1633571

```

**SOURCE** Kaposi's sarcoma-associated herpesvirus - Human herpesvirus 8.

ORGANISM Kaposi's sarcoma-associated herpesvirus  
Viruses: dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Rhadinovirus.  
REFERENCE 1 (bases 1 to 3489)

TITLE	MOORE, P. S., CHANG, Y. AND KNOWLES, D. M.
Kaposi's sarcoma-associated herpesvirus contains G protein-coupled receptor and Gv12 in D homolologs which are expressed in Kaposi's	

JOURNAL  
J. Virol. 70 (11), 8218-8223 (1996)  
MEDLINE  
97048116

TIME	AUTHORS
Direct Substitution	Bohenzky, R.A., Russo, J.J., Edelman, I.S., Chang, Y., Moore, P.S., Cesarman, E. and Knowles, D.M.

JOURNAL  
Submitted (21-MAR-1990) ROY A. BORNHATZKY, Pathology, Columbia  
University College of Physicians and Surgeons, 630 W. 168th Street,  
New York, NY 10032, USA

```
source
1. .3489
/organism="Kaposi's sarcoma-associated herpesvirus"
/organism host="Homo sapiens"
```

```

/clone="L54"
/cell_line="BC-1, a body-cavity based lymphoma cell line"
1..3489
CDS

```

```

/codon_start=1
/protein_id="AAC55944.1"

```

translation = "MAPGMRILRSGSTGATRGSKRKNRSPERCIDGDDLLHOPR  
RKHVADIDGREGCPHTLPFGSPYTVTSGLPAFVSSPTLPVAPISPEAPATPLPPLA  
LLPVTSSSPIPEHPVSPQTDTTSSSPALPTPOESSORPLSSPTGRPBDSPT

SPPVSPPRCDFAINRSVYPPWATESPIYVGSSEDDTTPRQPTSPISISSSPSEGSW  
 GDDTAMLYLAEIAEASKNEKECSENNQAGBNDGNDNISKEQYVDKDDNDNKKDBEE  
 GDDTAMLYLAEIAEASKNEKECSENNQAGBNDGNDNISKEQYVDKDDNDNKKDBEE

[illegible][illegible][illegible]

DESGDQFRRFLKGRDRPCQWKFPAVIFWMDNGLTGLKLSQAFQFGGKAGVPSCL  
ARQHIIFYRRFLKGRDRPCQWKFPAVIFWMDNGLTGLKLSQAFQFGGKAGVPSCL  
PHGPGDQSPIITYCYVVCQNKDTSKVKQMARLAWAESHPLAGNQLSSIVKFKKPLPLT

BASE COUNT ORIGIN	1053 a	862 c	1137 g	437 t
----------------------	--------	-------	--------	-------

Query Match	3.18;	Score 68.4;	DB 16;	Length 3489;
Best Local Similarity	47.68;	Pred. No. 1.8e-05;		

QY     979   gagctgttcatgacagaccgaggcgctggcagaagcgcggcagcgtgaactgcagcgg   1038  
|||||  
|||||

1039	caggagcttctcatgcagaagcgcgtctgcgatgcagtcaccaacagatccctccagaagcag	1098
------	---	------

	CAGGACGAGATGAACAGGAGCAGCAGGAGGAGCAGGAGCAGCAGGAGGAGCAGGAGCAG	2301
Db	2242 CAGGACGAGATGAACAGGAGCAGCAGGAGGAGGAGCAGGAGCAGCAGGAGGAGCAGGAGCAG	2301
	1099 CAGGAGATTGGAGCCCGCCCAAAAGGAAAAAATTCGCCCAAGAAGTCTCATGAAAATGAG	1158
yv		

Db 2302 GAGTTAGAGCAGCAGGAGTTAGAGATCAGGAGCAGGAGTTAGAGGACGAGAG 2361

[illegible]

Db 2362 CAGGAGTTAGAGCAGCAGCAGAGATTAGAGCAGCAGCAGAGATTAGAGCAGCAG 2421  
||  
QY 1219 gaagaagactggggtccaagaagacgactcttgccttaaaacctaactgctgagta 1278  
||  
Db 2422 GAGCAGGAGTTAGAGCAGCAGCAGAGCTTAGAGCAGCAGCAGAGATTAGAGCAG 2481  
||  
QY 1279 caccagtaacccttcgcaagcaagtaatgatacaggagatggaaactgagctcgagcc 1338  
||  
Db 2482 CAGGAGCAGAGATTAGAGCAGCAGCAGAGATTAGAGCAGCAGAGAGTGGAAGACAA 2541  
||  
QY 1339 gcaagatgacctgagatggaggaagcagagagagcagagagcagattccggaatatgag 1398  
||  
Db 2542 GAGCAGGAGAGTGGAAGACCAAGCAGCAGCAGAGAGCAGAGAGCAGAGATTAGAGAGTGAGAG 2601  
||  
QY 1399 ga 1400  
||  
Db 2602 GA 2603

RESULT 12  
AR065852/c  
LOCUS AR065852 32207 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 20 from patent US 5849564.  
ACCESSION AR065852  
VERSION AR065852.1 GI:5996068  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 32207)  
AUTHORS Chang, Y., Bohenzky, R.A., Russo, J.J., Edelman, I.S. and Moore, P.S.  
TITLE Polypeptides from Kaposi's sarcoma-associated herpesvirus, DNA  
encoding same and uses thereof  
JOURNAL Patent: US 5849564-A 20 15-DEC-1998;  
FEATURES  
source 1. 32207  
/organism="unknown"  
BASE COUNT 7229 a 9156 c 8713 g 7109 t  
ORIGIN

Query Match 3.1%; Score 68.4; DB 5; Length 32207;  
Best Local Similarity 47.6%; Pred. No. 2e-05;  
Matches 201; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 979 gagcgcttcatgacagaccggagcggtcgcaagggcgcgagcggtgagctgcagcg 1038  
||  
Db 19815 GAGCAGCAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAG 19756  
||  
QY 1039 caggagcttcatgcaagaagcggtcgatgagatcccaacaagatctctccaagagcag 1098  
||  
Db 19755 CAGCAGCAGAGATGAGCAGCAGCAGCAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAG 19696  
||  
QY 1099 caggagatggagcgcaaaaggaagaatctgcccaagaagcagcagaggaagaatgag 1158  
||  
Db 19695 GAGTTAGAGAGCAGCAGCAGAGATTAGAGATCAGAGCAGCAGAGATTAGAGAGCAGCAG 19636  
||  
QY 1159 agatcccgaaagagatggaacagattgtagggaggaaggaagaagttaagaagcaatg 1218  
||  
Db 19635 CAGGAGTTAGAGCAGCAGCAGAGATTAGAGAGCAGCAGCAGCAGAGATTAGAGAGCAG 19576  
||  
QY 1219 gaagaagactgggggtccaagaagacgactactcttgccttaaaacctaactgctgagta 1278  
||  
Db 19575 GAGCAGGAGATTAGAGCAGCAGCAGAGAGATTAGAGCAGCAGCAGAGAGATTAGAGAG 19516  
||  
QY 1279 caccagtaacccttcgcaagcaagtaatgatacaggagatggaaactgagctcgagcc 1338  
||  
Db 19515 CAGGAGCAGAGATTAGAGCAGCAGCAGAGATTAGAGAGCAGCAGAGTGGAAGACCAA 19456  
||  
QY 1339 gcaagatgacctgagatggaggaagcagagagagcagagagcagattccggaatatgag 1398  
||  
Db 19455 GAGCAGGAGAGTGGAAGACCAAGCAGCAGCAGAGAGAGCAGAGAGATTAGAGAGTGAGAG 19396  
||

QY 1399 ga 1400  
||  
Db 19395 GA 19394

RESULT 13  
KSU75698/c  
LOCUS KSU75698 137508 bp DNA VRL 03-MAY-1997  
DEFINITION Kaposi's sarcoma-associated herpesvirus long unique region, 80  
putative ORF's and kaposin gene, complete cds.  
ACCESSION U75698  
VERSION U75698.1 GI:2065526  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE

JOURNAL Kaposi's sarcoma  
MEDLINE Science 266 (3192), 1865-1869 (1994)  
REFERENCE 95090463  
AUTHORS 2 (bases 35021 to 55726)  
Moore, P.S., Gao, S.J., Dominguez, G., Cesarman, E., Lungu, O.,  
Knowles, D.M., Garber, R., Pellet, P.E., McGeoch, D.J. and Chang, Y.  
Primary characterization of a herpesvirus agent associated with  
Kaposi's sarcoma  
J. Virol. 70 (1), 549-558 (1996)  
3 (bases 28661 to 29741; 117919 to 118101)  
Zhong, W., Wang, H., Herndler, B. and Ganem, D.  
Restricted expression of Kaposi sarcoma-associated herpesvirus  
(human herpesvirus 8) genes in Kaposi sarcoma  
Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6641-6646 (1996)  
96270595  
4 (bases 122794 to 123567; 123809 to 127297; 129372 to 130400;  
130551 to 134441)  
Cesarman, E., Nador, R.G., Bal, F., Bohenzky, R.A., Russo, J.J.,  
Moore, P.S., Chang, Y. and Knowles, D.M.  
Kaposi's sarcoma-associated herpesvirus contains G protein-coupled  
receptor and cyclin D homologs which are expressed in Kaposi's  
sarcoma and malignant lymphoma  
J. Virol. 70 (11), 8218-8223 (1996)  
97048116  
5 (bases 1 to 137508)  
Moore, P.S., Boshoff, C., Weiss, R.A. and Chang, Y.  
Molecular mimicry of human cytokine and cytokine response pathway  
genes by KSHV  
Science 274 (5293), 1739-1744 (1996)  
97094384  
6 (bases 1 to 137508)  
Russo, J.J., Bohenzky, R.A., Chien, M.-C., Chen, J., Yan, M.,  
Maddalena, D., Parry, J.P., Peruzzi, D., Edelman, I.S., Chang, Y. and  
Moore, P.S.  
Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus  
(HHV8)  
Proc. Natl. Acad. Sci. U.S.A. 93 (25), 14862-14867 (1996)  
97121480  
7 (bases 1 to 137508)  
Russo, J.J., Bohenzky, R.A., Chien, M.-C., Chen, J., Yan, M.,  
Maddalena, D., Parry, J.P., Peruzzi, D., Edelman, I.S., Chang, Y. and  
Moore, P.S.  
Direct Submission  
Submitted (17-Oct-1996) Dept of Pathology/Genome Center/Division of  
Epidemiology, Columbia University, 650 West 168th Street, New York,  
NY 10032, USA  
8 (bases 1 to 137508)  
Russo, J.J., Bohenzky, R.A., Chien, M.-C., Chen, J., Yan, M.,  
Maddalena, D., Parry, J.P., Peruzzi, D., Edelman, I.S., Chang, Y. and  
Moore, P.S.  
Direct Submission

TITLE

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS

TITLE

JOURNAL  
MEDLINE  
REFERENCE

TITLE

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS

TITLE

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS

TITLE

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS

TITLE

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS

TITLE

TITLE





QY	1099	caaggatgagcagcgaagaaagaaattgcgccagaagcagagaggaaatgag	1158
DB	124996	GAGTTAGAGGAGCAGACGAGAGTTAGGAGATCAGAGACGAGACTTGAAGAGCAGGAG	124937
QY	1159	agataccggaagagatgtaagacaagatctgtaagagaggaagaatttaagaagaatg	1218
DB	124936	CAGGAGTTAGAGAGCAGACGAGAGTTAGAGAGCAGAGCAGAGAGTTCAGAGAGCAG	124877
QY	1219	gaagaagactggggtctcaagaagacagctactcttgcctaacaacatacctgctgagta	1278
DB	124876	GAGCAGAGGATTAGAGAGAGCAGAGAGTTAGAGAGCAGACAGAGTTCAGAGAGCAG	124817
QY	1279	caccacaagtaaccccttcgcaagcacaagatgatacaggagtgaggaacctgagctgagccc	1338
DB	124816	CAGGAGCAGAGAGTTAGAGAGCAGAGCAGAGTTAGAGAGCAGAGAGTTCAGAGAGCAG	124757
QY	1339	gcaagatgccttgatgagagcagcagagagcagagagcagagcttcggaagaatgag	1398
DB	124756	GAGCAGAGAGGTTGAGAGCAGAGCAGAGCAGAGAGCAGAGAGTTCAGAGAGGTGAG	124697
QY	1399	ga 1400	
DB	124696	GA 124695	
RESULT	14		
LOCUS	U93872/c		
DEFINITION	U93872 133661 bp DNA VRL 09-JUL-1997		
KEYWORDS	Kaposi's sarcoma-associated herpesvirus glycoprotein M, DNA replication protein, glycoprotein, DNA replication protein, FLICE inhibitory protein and v-cyclin genes, complete cds, and tegument protein gene, partial cds.		
ACCESSION	U93872		
VERSION	U93872.1	GI:2246466	
ORGANISM	Kaposi's sarcoma-associated herpesvirus - Human herpesvirus 8. Kaposi's sarcoma-associated herpesvirus		
TITLE	Virus; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Rhadinovirus.		
REFERENCE	1 (bases 29032 to 30108; 117733 to 118431)		
AUTHORS	Zhong, W., Wang, H., Herndler, B. and Ganem, D.		
TITLE	Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6641-6646 (1996)		
MEDLINE	96270595		
REFERENCE	2 (bases 17242 to 17856)		
AUTHORS	Neipel, F., Albrecht, J. C., Ensser, A., Huang, Y. Q., Li, J. J., Friedman-Kien, A. E. and Fleckenstein, B.		
TITLE	Human herpesvirus 8 encodes a homolog of interleukin-6		
JOURNAL	J. Virol. 71 (1), 839-842 (1997)		
MEDLINE	97138401		
REFERENCE	3 (bases 123309 to 124082)		
AUTHORS	Li, M., Lee, H., Yoon, D. W., Albrecht, J. C., Fleckenstein, B., Neipel, F. and Jung, J. U.		
TITLE	Kaposi's sarcoma-associated herpesvirus encodes a functional cyclin		
JOURNAL	J. Virol. 71 (3), 1984-1991 (1997)		
MEDLINE	97184528		
REFERENCE	4 (bases 122660 to 12326)		
AUTHORS	Thome, M., Schneider, P., Hofmann, K., Fickenscher, H., Mehl, E., Neipel, F., Matmann, C., Burns, K., Bodmer, J. L., Schrotter, M., Scalfidi, C., Krammer, P. H., Peter, M. E. and Tschopp, J.		
TITLE	Viral FLICE-inhibitory proteins (FLIPs) prevent apoptosis induced by death receptors		
JOURNAL	Nature 386 (6624), 517-521 (1997)		
MEDLINE	97242415		
REFERENCE	5 (bases 1 to 133661)		
AUTHORS	Neipel, F., Albrecht, J. C. and Fleckenstein, B.		
TITLE	Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus human herpesvirus 8: determinants of its pathogenicity?		
JOURNAL	J. Virol. 71 (6), 4187-4192 (1997)		
MEDLINE	97296220		

REFERENCE	6 (bases 1 to 133661)		
AUTHORS	Neipel, F., Albrecht, J. C., Ensser, A., Huang, Y. Q., Li, J. J., Friedman-Kien, A. E. and Fleckenstein, B.		
TITLE	The genome of human herpesvirus 8 cloned from Kaposi's sarcoma		
JOURNAL	Unpublished		
REFERENCE	7 (bases 1 to 133661)		
AUTHORS	Neipel, F., Albrecht, J. C., Ensser, A., Huang, Y. Q., Li, J. J., Friedman-Kien, A. E. and Fleckenstein, B.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-MAR-1997) Virology, University of Erlangen, Schlosgarten 4, Institut für Klinische und Molekulare, Erlangen 91054, Germany		
FEATURES			
SOURCE	Location/Qualifiers		
	1. 133661		
	/organism="Kaposi's sarcoma-associated herpesvirus"		
	/db_xref="taxon:37296"		
	/note="sequence derived from overlapping Lambda phage clones isolated from two libraries of Sa3a partially digested DNA from Kaposi's sarcoma biopsies"		
	105. 959		
	/note="ORF K1"		
	/codon_start=1		
	/protein_id="AAB62619.1"		
	/db_xref="GI:2246494"		
	/translation="MPEYVCSLAVERGLLSLSQSPMLCGVISTPTTICPSNT SLPTSWICNDTRLRLTQOTFTVDTLLICNSCVGSGHRSILMTWTPQVLDITLQD PSNTVTCGAVHTLYCSTGNNVYVHMLPNRNFNEVSQTATNFNLNQGCVACSG LSLSRLNRLCFASRACANLTPETHFVSLSSPTTGTSTNRRGTTLDLVKREKSTLL HIOVHLFVEMTVALIGTMCIGITTFIFACQQRQDSKNTVPQDLOYVSLDCTED YVQDVWY"		
	1127. 2779		
	/note="ORF 04, complement control protein homolog"		
	/codon_start=1		
	/protein_id="AAB62602.1"		
	/db_xref="GI:2246477"		
	/translation="MARIROPTIMTFWVAVIODNKKCSOKTLIGRLKMSDQDA VGEVVELRCSGTTTYARNTATTCLOGSTSEPTATCNKSCSNRPETLQKYLPHGQ ODALKTGANISYVCSNEGYFLVGRVYVAVICGASQWAMSSPFCCKECHKPKLEN GDRKPDVYEVDAVHEFECNEGTYLVGPHSICAVNNWTSMPTCELAGCEFSPT HGYPIDGFSLTLYHKQSVTPACNDGFEVLRSPITCTCAVTEMDPEPLKCVLEDDISV NNGRLPLFNEKPNQGVFORSVNTEPTETPCVTDCTNCEOPKLLPTSEGFN ETTTSSNTITQODEKETSQPNHTITSAISMKAKGNTTSEKATISTYSHSNKMDTGTI DAPRSPISVQTPVNTNNAPTRTLSLHIEGSPNSITSEKATISTYSHSNKMDTGTI YTLNKTQTPSPNKNPNSQAKSSTKRVETNKTYSNPASTLSADVPQRRPEPTL PPIFRPPASKNRYLEKOLVIGLTAVALMGLTFLHFLPFR"		
	3194. 6592		
	/note="ORF 06, major ssDNA binding protein homolog"		
	/codon_start=1		
	/protein_id="AAB62603.1"		
	/db_xref="GI:2246478"		
	/translation="MALKGPQTEENIGSAAAPGCGYLAVYLTHNFPIDEAALNG YPAKVFSLPLGLHVEDSPFLVAKVHKIKIDATTAASVLTYSYHRAIFAHNTLFO PIRQGLGELKRESERELPGESMFLVBOHKGTLMSAPOLCPACIDPMAYVITGEG KEHLGKGLVLPVSQTTPVHIGHQAFKILYDEDLFGSRADQLCFRTNPDISRIYL DSIFTGQAQLRKNDSTVYIQASERQVHQVYIKPLQAKDRPQASRGTDSLTLY IDSLVLELHSGYSLPTFEGPODCEVLYNDVTPITFECENPDRALRLAEWAEOQLH IGADLPANSVLILTRAKLPKONKODANMYNSFYLOHGLGLSATVAYEASGSAFK KIVHOSMFCDLPQDAIGMPAPARQVYIAAMLVPTIKIRIKIIFNSGTGES IOAGFMKASORSTYVCGPYMKFLNAHKTLPSPKTSALYLMHRTIGOTPKYPIIG VSGEHLTELCNYKASQAFEEINVDLYVDITLTSTAKIKLNSILRACQOTFYATT LSLSP		

CDS	
	note="ORF 7, processing and transport protein homolog (ICP18-5)"



gene /db\_xref="taxon:37296"  
/note="right end of the unique region plus 8 complete  
copies and 1 partial copy of the terminal repeat (TR)"  
1..2105  
/gene="ORF 68"  
1..1102  
/gene="ORF 68"  
/function="involved in DNA packaging"  
/note="similar to HSV-1 UL32, HCMV UL52 and EBV BFLF1"  
/codon\_start=2  
/evidence=not\_experimental  
/protein\_id="AAD46496.1"  
/db\_xref="GI:5669889"  
/translation="IMHLFPEEEATYGLFGPRLMGIDLOLHFFVOKCFKTTAAEKI  
IGISNLOFLSEFIRGMLTGTTTCNCFKSTMPRTDKEATGPTCCOITDTTAPAS  
GIPELARATFGASRPCKPSILPALIDIMSTSSLEDEPRRLISDMSLEKSVASH  
DPFSPRIQADTOSGRCMLMPTLGRKNGTASVCLCECLAAHPBARALDLOCEV  
MHIENNVKLVDRIAFVLDNPPAMPYVSOPFLRELIRGCTPOEIKHLFCDFCALNA  
KVSVSDVLEFLRPEQDYKRLKRSAAAGOLDANTLFDEEVOTLVFLFKGLONARVKG  
TTSLDIIRLEITFQOLKRHRIDLNAHPSOTSLYA"  
1135..2105  
/gene="ORF 69"  
1135..2043  
/gene="ORF 69"  
/note="similar to HSV-1 UL31, HCMV UL53 and EBV BFLF2"  
/codon\_start=1  
/evidence=not\_experimental  
/protein\_id="AAD46497.1"  
/db\_xref="GI:5669890"  
/translation="MPKSYSSHISLATSTGNSGPPDIRCLSSRLRSVPPGARSAVS  
SKHRNGLRKFIISDKVFEILSHRHELGVDFLEKEMPTICTSKTVMPLDLSTVAGRC  
VLSHPFGHSSNMGFQALCPSTENPTVAQSGRPQTMVGDLAKKNELCSVALAFYHHA  
DKVHKTFEYLSLHSMQDVYKSFLOPGLIYALVTKFEGHPLPFTTNGMLTMC  
ILFKRALHGETALRLMDNLPYKISADCCRSQYVVKFVPTHPTASIAVQVHTIC  
EAAVALDCTDEKRDIDQKGTALVYNA"  
2100..2105  
polyA\_signal  
/gene="ORF 69"  
/note="possible signal for ORFs 68 and 69"  
/evidence=not\_experimental  
complement(2126)  
polyA\_site  
/gene="ORF K12"  
/evidence=experimental  
complement(2126..2824)  
/gene="ORF K12"  
/product="T0.7"  
/evidence=experimental  
complement(2126..2824)  
gene  
/gene="ORF K12"  
complement(2139..2144)  
/gene="ORF K12"  
/evidence=experimental  
complement(2616..2798)  
/gene="ORF K12"  
/note="transforming gene"  
/codon\_start=1  
/evidence=experimental  
/product="Kaposin"  
/protein\_id="AAD46499.1"  
/db\_xref="GI:5669892"  
/translation="MDRGITVFVAHVVDVLLNGMRMLGALPPLVCLLAISVPPSG  
ORGPVAFRTVATGAN"  
5'UTR  
complement(2799..2824)  
/gene="ORF K12"  
/evidence=experimental  
2820..3081  
repeat\_region  
/rpt\_type=direct  
/rpt\_unit=2820..2842  
3082..3436  
repeat\_region  
/rpt\_type=direct  
/rpt\_unit=3084..3106  
complement(6928..6933)  
polyA\_signal  
/gene="ORF 71"  
/note="possible signal for ORF 71"

gene /evidence=not\_experimental  
complement(6928..7550)  
/gene="ORF 71"  
complement(6984..7550)  
/gene="ORF 71"  
/function="inhibitor of apoptosis"  
/note="contains two death effector domains"  
/codon\_start=1  
/evidence=experimental  
/product="v-FliP"  
/protein\_id="AAD46498.1"  
/db\_xref="GI:5669891"  
/translation="MATYEVLCVARKIGTQDREVVLFLNFIPOPTLAQIGALRA  
LKEGRLPFLAECLFAGSRDLRLDLHDPRLERLHAGTMSYFSPQUTVLAVD  
GELCARDLRSILFSLKDTIGSRSTPOTFLHMYCWMENLDLGPFDVALMSLRLSLR  
VDLQOVOTLMKIHLSGSHSQHYHTTP"  
complement(7601..7606)  
polyA\_signal  
/gene="ORF 72"  
/note="possible signal for ORFs 72 and 73"  
/evidence=not\_experimental  
complement(7601..8406)  
/gene="ORF 72"  
complement(7601..12037)  
/gene="ORF 73"  
complement(7633..8406)  
/gene="ORF 72"  
/note="cyclin D homologue"  
/codon\_start=1  
/evidence=experimental  
/product="v-cyclin"  
/protein\_id="AAD46500.1"  
/db\_xref="GI:5669893"  
/translation="MAPANNPSSGLDPTICEBRLFYVNIIEPFLTSDVSFGEFO  
SLTSHMRKILGFMFSGCOEYVLRBNVVALNLDRLILLKOVSEHROKTSACIL  
VASKRLSTPSTSSLCYAAADSFSGEILIDEXELKEKLRATVATVATVTSLLL  
KILGSSQHLDEWHHEVNTLITKALVDPTGSPASIIAAGCALLVANVLPDTHSG  
GVVPLQASILIGDVSVLQAAVEQILITSVSDFDLRLDSY"  
complement(8648..12037)  
CDS  
/gene="ORF 73"  
/note="LNA: component of latency-associated nuclear  
antigen, LANA"  
/codon\_start=1  
/evidence=experimental  
/product="latent nuclear antigen"  
/protein\_id="AAD46501.1"  
/db\_xref="GI:5669894"  
/translation="MAPPGMRILRSRGTGAPITRSCSKRRNRSPPRCULGDDLIQPR  
RKHVADSVGRKCGPHTLPIGSPVFTVSGLPFAVSPPLPAPIPSPAPPLPAPA  
LLPVTSSSPPIPSHPVSPGTTDHPSPALPYPQSESSQRPPLSSPFGPDSSTP  
MKPPSQOTTPHSPPTPPPEPSKSPDSLAPLRLSLRLKRLSSPQGPSTLANICQ  
SPVSPRCDFANRSVPPMATESPIYVGSSSDGTPPROPTSPISIGSSPSBSGM  
GDDTAMLVILAEILAEASKNEKESENNAGDNGDNITSKESQVDRKDDNNKDEEE  
DETDDE  
EDDDDE  
PQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQ  
REPQOREPQOREPQOREPQOREPQOREPQOREPQOREPQOREPQOREPQOREPQO  
QOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQ  
DEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQO  
QOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQ  
QOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQ  
EQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQ  
EQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQ  
QLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQO  
MEVDYPVVSTHEQILASSPQGNTPDDDPQPGSRRYRVLRTSPHRRGVARRVYPT  
HPKRPHPYQOQPPVYRQIDCCPAKARQHI FYREFLKGDKGRDCKCOMKRAVITWGN  
DPTGLKTLQSAHQEGSVAKAGVPSCLPHFEGPDOSPITTYVYVYQNKDTSKRVQARLA  
WEASHPLAGNLQSSIVKPKPLPPLPQEPNGDPSQDEM"  
complement(9375..11047)  
repeat\_region  
/note="complex array of repeats"  
/rpt\_type=direct  
12624..15262  
/gene="ORF K14"  
12624..13670  
/gene="ORF K14"

```
/note="similar to OX-2 membrane glycoprotein"  
/codon_start=1  
/evidence=not_experimental  
/protein_id="AAD46502.1"  
/db_xref="GI:5669895"  
/translation="MIHTEFDCPPRRVEGVYSSYFLGAPGRTAIKTEGVSA  
LNPVTVAVAGTGVSVRVVVCAPMTTSPASRGSSLSFLSPVAFILALGAVGA  
RVGGPMRGSAAALCAITPRADIVSYMOKROLPGPVNATVSHSGVYVOTQYRHRAN  
IYCPGLMNSTVLYHNLAVDDEGCYLCIFNSFGCRQVSCIRACLEVTSPPTGHVYNSTE  
DADVTCLATGRPPPVNTWMAAPWNHNSSTOEPTSDGLVAMRTVRLPRGDNTPSE  
GICLITWGNESISIPASIOGLPLADLPAAGTLAGVAITLVGLFGLFALHCRKQGG  
ASPTSDMDPLSTQ"  
1411..15262  
/gene="ORF 74"  
1411..15139  
/gene="ORF 74"  
/note="G protein coupled receptor; similar to IL-8  
receptor; transforming protein"  
/codon_start=1  
/evidence=experimental  
/product="v-gPCR"  
/protein_id="AAD46503.1"  
/db_xref="GI:5669896"  
/translation="MAADEFLTFILDDDESNETLNMSGYDSGNFSLEYVCEMTTV  
VPYTWVGLISLIFLIVNGLVITIFCKHRSRAGADIDILIGLINSICLSISILA  
EVLMLFPNISTIGLCRLIEFFLYLYVIDIFSVVCSLVRYLLVAVSTRMPKROSL
```

Query Match 3.0%; Score 67; DB 16; Length 28559;

Best Local Similarity 47.3%; Pred. No. 4.2e-05; Mismatches 225; Indels 0; Gaps 0;

```
QY 976 cggagagctgttcataagacagccggagcgctgcaagagcgcgcaagctcctcagag 1035  
DB 9847 CAGGAGTTAAGAGCAGAGCAGAGAGCTTAGAGAGCAGCAGAGCTTAGAGAGCAG 9788  
QY 1036 cggcagagcttcctcagcgaagcgctgcaagagctcctcagag 1095  
DB 9787 GAGCAGAGCTTAGAGAGCAGAGCAGAGCTTAGAGAGCAGAGCAGAGCTTAGAGAG 9728  
QY 1096 cagcagagagatygagcgcaagagagaaatgcccagaagcagcagagagaaat 1155  
DB 9727 CAGGAGCAGAGCTTAGAGAGCAGAGCAGAGCTTAGAGAGCAGAGCAGAGCTTAGAG 9668  
QY 1156 gagagataccggaagagatygaaacagatgtgaagagagagaaatgtaagaaacaa 1215  
DB 9667 GAGCAGAGCAGAGCTTAGAGAGCAGAGCAGAGCTTAGAGAGCAGAGCAGAGCTTA 9608  
QY 1216 tgggaagaagactlggagctcaagaaagactactcttgccctaaaccatcactgag 1275  
DB 9607 GAGGAGCAGAGCAGAGCTTAGAGAGCAGAGCAGAGCTTAGAGAGCAGAGCTGGA 9548  
QY 1276 gtacacccaataccctctcgaagccaagtatgatcaggagtggaacctgagctcag 1335  
DB 9547 GAGCAAGAGCAGAGCAGAGCAGAGCTTAGAGAGCTGAGAGCAGAGCAGAGAG 9488  
QY 1336 ccgcagagatgactgagtgagagcagagagagagagagagatccggaaatcat 1395  
DB 9487 CAGGAGAGCAGAGCAGAGCAGAGCTTAGAGAGCTGAGAGCAGAGCAGAGCAGAGCTTA 9428  
QY 1396 gaggaag 1402  
DB 9427 GAGGAGG 9421
```

Search completed: May 20, 2000, 01:07:30  
Job time: 14654 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 19, 2000, 21:03:16 ; Search time 6491.04 Seconds  
(without alignments)  
-232.593 Million cell updates/sec

Title: US-09-502-945-1  
Perfect score: 1552  
Sequence: 1 cttctgatagcatacgcagaaa.....aaatgacttttaagaaga 1552

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 segs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_com:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_com:\*  
22: em\_ov:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pr5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1552	100.0	1552	40 AF039690	AF039690 Homo sapi
2	504	32.5	1325	51 AF161348	AF161348 Homo sapi
3	326.4	21.0	1198	11 HS079271	HS079271 Human clone
4	257.2	16.6	576	13 G27428	G27428 human SFS S
5	90.6	5.8	7218	5 I66494	I66494 Sequence 14
6	67.2	4.3	1686	24 E08995	E08995 DNA encodin
7	65.8	4.2	180385	40 AC007461	AC007461 Homo sapi
8	62.4	4.0	39103	35 CEI03368	CEI03368 Caenorhabd
9	62.4	4.0	193188	41 AC006884	AC006884 Caenorhab
10	59	3.8	17522	12 M05MHC4H2S	M17440 Mus musculu
11	55.6	3.6	2554	7 SCYK1202W	Z28201 S.cerevisia
12	55.6	3.6	3997	5 E14049	E14049 gDNA which
13	55.4	3.6	5420	34 AF056936	AF056936 Plasmodi
14	54	3.5	4000	7 D83006	D83006 Saccharomyc
15	54	3.5	100000	9 AP000067	AP000067 Homo sapi
16	53.2	3.4	37797	34 CET21B4	Z81124 Caenorhabd
17	53.2	3.4	69134	32 CEY57A10_4	Continuation (5 of
18	53.2	3.4	110000	32 CEY38E10_1	Continuation (2 of
19	52.8	3.4	110000	32 PFMA14P1_0	AL034557 Plasmodiu
20	52.8	3.4	148278	40 AC006101	AC006101 citb_338_
21	52.6	3.4	175630	56 AC023059	AC023059 Mus muscu
22	52.4	3.4	20776	34 CEK03D10	Z81561 Caenorhabd
23	52.4	3.4	110000	32 CEY63D3_0	Continuation (2 of
24	52.4	3.4	110000	32 CEY63D3_1	Continuation (2 of
25	52.2	3.4	22502	41 AC006832	AC006832 Caenorhab
26	51.8	3.3	120000	41 AC007287	AC007287 Homo sapi
27	51.8	3.3	151428	55 AC016063	AC016063 Homo sapi
28	51.6	3.3	4238	12 RNCNG41	AJ000496 Rattus no
29	51.6	3.3	315277	43 AC018634	AC018634 Homo sapi
30	51.4	3.3	149886	12 MMHC438N12	AF049850 Mus muscu
31	50.8	3.3	879	13 CNS01JRG	AL147405 Anopheles
32	50.4	3.2	720	9 AB007820	AB007820 Homo sapi
33	50.4	3.2	1410	9 AB009843	AB009843 Homo sapi
34	50.2	3.2	155364	52 AC022248	AC022248 Homo sapi
35	50	3.2	227070	56 AC023888	AC023888 Homo sapi
36	49.6	3.2	1512	9 HSKERELP	X62571 H. sapiens m
37	49.4	3.2	3987	43 AC017190	AC017190 Drosophila
38	49.2	3.2	3371	4 GGCLAUSA	X67778 G.gallus mR
39	49.2	3.2	256172	41 AC005139	AC005139 Plasmodiu
40	49.2	3.2	308399	44 AC005140	AC005140 Plasmodiu
41	49	3.2	116515	43 AC013527	AC013527 Homo sapi
42	49	3.2	173690	43 AC011307	AC011307 Homo sapi
43	49	3.2	178975	41 AC011302	AC011302 Homo sapi
44	48.8	3.2	265640	57 AC008046	AC008046 Homo sapi
45	48.8	3.1	64411	32 CEY63D3_3	Continuation (4 of

ALIGNMENTS

RESULT 1  
AF039690 1552 bp mRNA PRI 05-MAY-1999  
DEFINITION Homo sapiens antigen NY-CO-8 (NY-CO-8) mRNA, partial cds.  
ACCESSION AF039690  
VERSION AF039690.1 GI:3170179  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Scanlan,M.J., Chen,Y.T., Williamson,B., Gure,A.O., Stockert,E.,  
Gordan,J.D., Turel,O., Sahin,U., Pfrendrich,M., and Old,L.J.  
TITLE Characterization of human colon cancer antigens recognized by  
antibodies  
JOURNAL Int.J. Cancer 76 (5), 652-658 (1998)  
MEDLINE 98272252  
REFERENCE 2 (bases 1 to 1552)  
AUTHORS Scanlan,M.J.  
TITLE Direct Submission  
JOURNAL Submitted (23-DEC-1997) Ludwig Institute for Cancer Research, New  
York Branch at Memorial Sloan-Kettering Cancer Center, 1275 York  
Avenue, New York, NY 10021, USA  
FEATURES  
source  
1. 1552  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="colorectal carcinoma"  
/gene="NY-CO-8"  
/gene="NY-CO-8"  
/gene="NY-CO-8"  
/codon\_start=1  
/product="antigen NY-CO-8"  
/protein\_id="AAC18039.1"  
/translation="LIDASEKLTLYEKECEIESOLKFLRNDLAEYORFCEDLKBL  
KKEFLIANTGNRVGLCLKCAOHEAVLSOHTVNHOTIELEVERDMLALYSV  
RSSIADTPOORBAAYEOKVOIOISEANPEKRALIODDLKELERDAERLEKELA  
SOEKRAITEKDMKKEITKEREKMSKMLISNTIQLAQVEKYTKETISATNOEE  
IOSJLAKRMDYTKVCGEMRYOLANKTMEKDEKEREKREKAKTNDLEKDOELK  
RIELDESKEHLQEOQKALAREECLRLTELLGESHOHLRSELAIOSQEKRYTD  
KIGKLORNNELEEOCVOHGRVHEITMKORLRLDKHSQATADQLVOLSKONDLIER  
OSLSEVDRLRQLPSPQSDC"  
BASE COUNT 525 a 291 c 401 g 335 t  
ORIGIN  
Query Match 100.0%; Score 1552; DB 40; Length 1552;  
Best local Similarity 100.0%; Pred. No. 0;  
Matches 1552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ctctctgatagcacgcaggaagctaaacctactatgatgagaaagtgtgaattgagaa 60  
DB 1 CTCTGTGATGCATCCGAGAGCTAAACTTACTTATGAGAAAAGTGTGAATTGAGAA 60  
QY 61 tcccaattgaagtcttttggagaacgctagctgaatatagaagaacttgtgaagcttc 120  
DB 61 TCCCAATTGAAGTGTGAGGAAGAGACTTAGCTGAATATATGAGAACTTGTGAAGATCTT 120  
QY 121 aaagagaactaagaactaagaattctctctgctgctataactgttaaccgttgtgt 180  
DB 121 AAAGAGCACTAAAGCTTAAGAAATTTCTTCTGCTCTTAATCTGTAAACCGTGTGGT 180  
QY 181 ggtcttgtttgaatgtgtctcagcatgaagcgtgtctctcccaaaccaactaatgt 240  
DB 181 GGCTTGTGTTGAATGTGCTCAGCATGAAGCTGTCTTCCCAAAACCACTTAAATGTT 240  
QY 241 catatgcagacatctgaaagactggttaagaagaagatgactgtatgtctgactagt 300  
DB 241 CATATGCAGACATCTGAAAGACTGTGTTAAAGAAAGATGATCTTGATGTGCACTAGTT 300

QY 301 tccgtaagagcagcttgcgagatacgcagcaagaagaagcaagtgtctta tgaaccagt 360  
DB 301 TCCGTAAGAGCAGCTTGGCAGATACGCACAGAAAGAGCAAGCAAGTCTTATGAACAGAGTG 360  
QY 361 aaacaagcttctgcaatatctgaggaagccaatttgaaaaaaaccaagctttaaccag 420  
DB 361 AAACAAGCTTCTGCAAAATCTGAGAGGAGCAATTTTAAAAAACCAAGCGCTTTAATCTCAG 420  
QY 421 tgtgaccatttgaaggaagccttgagagcagcagcagcagcacttgaagaactgtca 480  
DB 421 TGTGACCATTGAGGAAGAGAGCTTGAGACAGCGAGCGACTTGAAGAAAGAACTTGCA 480  
QY 481 tctcagaagaagaagaagccattgagaagaacatga tgaaaaaaggaataaagaagaa 540  
DB 481 TCTCAGCAAGAGAAAGGCGCATTGAGAAAGACATGATGAAAAAGAAATTAACGAAGAA 540  
QY 541 agggagtaacatggtatcaaaagtgtatcttgtctcagaatatctcccaactgtaggc 600  
DB 541 AGGGAGTACATGGGATCAAAAGATGTTGATCTGTCTCAGAAATATTTGCCAACTGGAGGCC 600  
QY 601 caggtgaaaaagtttaaaagaaagatttcagctataactgagaggaattcaa 660  
DB 601 CAGGTGAAAAAGTTTACAAAGAAAGATTTCAGCTATTATCAACTGAGGAGAAATTCAA 660  
QY 661 agcagctggtctctcgggaatgtagtcaaaaagtgtgtgtagaagaatgcctacag 720  
DB 661 AGCGAGTGGCTTCTCGGGAATGAGATGTCACAAAGTGTGTGAGAGAAATGCGGTATCAG 720  
QY 721 ctgaataaacaacacatgtgagaagatgtagcagaagaagaagcagagatcagaaga 780  
DB 721 CTGAATAAACAACATGAGGAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780  
QY 781 aaactcagaagagctcttgaataaagaatcagaagaataagaagaattgagaactg 840  
DB 781 AAACTACAGGGATCTTGAATTAAGATGAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 840  
QY 841 gatgaagaacaacaacacttgcagcagcagcagcagcagcagcagcagcagcagcagcag 900  
DB 841 GATGAAGAACAACAACCTTGGAAACAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900  
QY 901 tgcctgagactaacaagacgtgcgtggtggaatctgagacccaactgtcacctcacagatc 960  
DB 901 TGCTTGAGACTTAACAGAACTGCTGGGGAACTGTGAGCAACCACTGCACTTACCAGATCT 960  
QY 961 gaaatagctaacctcagtcagaagaaaaaggtatatacatatataattggaagaattcag 1020  
DB 961 GAAATAGCTTAACCTCAGTCAAGAAAAAGGTATATATATATTAATTGGAAAAATTACAG 1020  
QY 1021 agaagaatgaaagaatttgaagacagtggtccagcatgtggaagtacatgaagcagt 1080  
DB 1021 AGAAGAAATGAAGAAATTTGAGGAACAGTGTCTCAGCATGGAGAGTACATGAGAGCATG 1080  
QY 1081 aagcaaaagcttaagcagcgtcgtgataagcacagccagcagcagcagcagcagcagcag 1140  
DB 1081 AAGCAAAAGCTTAAGCGCAGCTGATTAACACAGCAGCCACAGCCAGCGCTGTGTGACAG 1140  
QY 1141 ctctcagaagcagaagcagcgtctctcctcgtgagaggaagagcgtgtcgtggaaggtgagc 1200  
DB 1141 CTCTCAGCAAGCAGAGACAGCTTCTCTCTGAGAGGACAGAGCCTGTGTGGAAGAGTGGAC 1200  
QY 1201 cggctgcgagccagcttaaccagcatgcccacaaacttgatgtcagccttgaatgagacaga 1260  
DB 1201 CGGCTGCGAGCCAGTGTACCCAGCATGCCCAACATTCGATGCTGATGATGGAACAGA 1260  
QY 1261 gtgaataaataatgaattacaagaagataattcaatcgaatcgtgttgacttaataatgcca 1320  
DB 1261 GTGAATAAATAATGAATTAACAAGAAGATATTTACATTCATCTGTTTGAAGACTTATATG 1320  
QY 1321 caacgacacagacacttccagaggtgaacacgcctcagcctgaagtgagggtctgtgctcta 1380  
DB 1321 CAACGACACAGACCTTCCAGAGGTGACACCGCTCAGCTCAGTGGAGGTGCTGCTCTCA 1380



```
Qy 1381 tcaacgcgagcgtctgccccgacgacgtcggcgtgagctgagctctactactagctga 1440
|||||
Db 1381 tcaacgcgagcgtctgccccgacgacgtcggcgtgagctgagctctactactagctga 1440
|||||
Qy 1441 gcaagcctcgtgtatgttctcagaanaatggtctgaagtatgttcttaaatctgctcat 1500
|||||
Db 1441 GCAGACTCCTGCTGATGTTTTCAGAAATGCGCTGGAAGTATGTGTAAATCTGCTCAT 1500
|||||
Qy 1501 tctgtatgtatgattatcatatgatttctcaataatgtaacttttaagaaga 1552
|||||
Db 1501 TCGTATGCTAGGTTATACATATGATTTTCAATTAATGAACCTTTTAAAGAAA 1552
|||||

RESULT 2
AF161348 1325 bp mRNA PRI 01-FEB-2000
LOCUS AF161348 Homo sapiens HSPC085 mRNA, partial cds.
DEFINITION AF161348
ACCESSION AF161348
VERSION AF161348.1 GI:6841109
KEYWORDS FLI.CDNA.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1325)
AUTHORS Zhang, Q.H., Ye, M., Zhou, J., Shen, Y., Wu, X.Y., Guan, Z.Q., Wang, L.,
Fan, H.Y., Mao, Y.F., Dai, M., Huang, Q.H., Chen, S.J. and Chen, Z.
Human partial CDS cloned from cd34+ stem cells
Unpublished
2 (bases 1 to 1325)
Zhang, Q.H., Ye, M., Zhou, J., Shen, Y., Wu, X.Y., Guan, Z.Q., Wang, L.,
Fan, H.Y., Mao, Y.F., Dai, M., Huang, Q.H., Chen, S.J. and Chen, Z.
Direct Submission
Submitted (14-MAY-1999) Shanghai Institute of Hematology, Shanghai
Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II,
Shanghai 200025, P. R. China
FEATURES
SOURCE Location/Qualifiers
1. 1325
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CBDAND10"
/cell_type="cd34+ stem cell"
/tissue_type="cord blood"
147..>1325
/codon_start=1
/product="HSPC085"
/protein_id="AAF28908.1"
/db_xref="GI:6841110"
/translation="MAKSPENSTLEELIGVORS/REHNSRSIHQTCALKEGVTTIG
EDAPNISFSTSVGNEADARTAPPELOQSNAVQLDLROQDKSEVSPSRKMSPL
RSLEHETNMTFMDLVHTINDOSYIHLEAEVFKCEELSGMKNKIOVVLENEGL
QOQLKSORQETLEQFLDASGNMHNSTTGDGSGEFSKRPFSHDNDFFKAAS
AGEQLELEKLTLYEKECEIEESQLEPLRDLAEYORCEDEKLEQLKEFLAANTC
NRVGLGLCKCAHEAVLSQHTNMVQCTIEFLVKEERDLMALVSVRSLSLDTQOREA
SAYQVQOVQIISEANFEKTKALIQCDQLKEERQAEERLEKELEASQOEKRALEKDM
M"

BASE COUNT 430 a 269 c 338 g 288 t
ORIGIN

Query Match 32.5%; Score 504; DB 51; Length 1325;
Best Local Similarity 100.0%; Pred. NO. 5e-106;
Matches 504; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Db 942 CATAAAGATTTCTTGCTGCTGCTATACTGTAAACCGTGTGGTCTTTGTTGAAA 1001
|||||
Qy 196 tgtgtcgaacatgaagctgtcttcccaaacccactactatgttcatatgacacatc 255
|||||
Db 1002 TGTGCTCAGACATGAAGCTGTCTTCCCAAAACCACTACTAATGTTCAATATGACAGACATC 1061
|||||
Qy 256 gaaagactgtgttaagaagaagatgactgtgtcgtactgtttccgtaaggagcagc 315
|||||
Db 1062 GAAAGACTGTGTTAAAGAAAGATGATGATGCTGCACTAGTTCCTGAAGAGACAC 1121
|||||
Qy 316 ttggcagatcgcagcaagaagaagaagctgtcttataagcaggtgaagaagctttgcaa 375
|||||
Db 1122 TTGGCAGATACCGACAGCAAGACAGCAAGTCTTATGAAACAGTGAACAAATTTTGCA 1181
|||||
Qy 376 atatctgaggaaagccaatttgaaaaaaaccaaggtcttaaccagtgtgaccagtgaag 435
|||||
Db 1182 ATATCTGAGAGAGCCAAATTTGAAAAACCAAGGCTTTATTCAGTGTGACAGTTGAGG 1241
|||||
Qy 436 aaggagctgtgaaagcagcgcgagcgttgaagaagaagctgtcatctcagaagaaga 495
|||||
Db 1242 AAGGAGCTGGAGAGCGAGCGGACCTTGAAAAAGAACTTGCAATTCAGCAAGAGAAA 1301
|||||
Qy 496 agggccattgagaagaacatgatg 519
|||||
Db 1302 AGGCCATTGACAAAGACATGATG 1325
|||||

RESULT 3
HSU79271/c 1198 bp mRNA PRI 25-MAR-1997
LOCUS HSU79271 Human clones 23920 and 23921 mRNA sequence.
DEFINITION U79271
ACCESSION U79271
VERSION U79271.1 GI:1710237
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1198)
AUTHORS Andersson, B., Wentland, M.A., Ricafrente, J.Y., Liu, W. and Gibbs, R.A.
A 'double adaptor' method for improved shotgun library construction
Anal. Biochem. 236 (1), 107-113 (1996)
JOURNAL 96207227
MEDLINE 2 (bases 1 to 1198)
REFERENCE Yu, W., Andersson, B., Morley, K.C., Muzny, D.M., Ding, Y., Liu, W.,
Ricafrente, J.Y., Wentland, M.A., Lennon, G. and Gibbs, R.A.
Large Scale Concatenation cDNA Sequencing
Unpublished
3 (bases 1 to 1198)
AUTHORS Yu, W. and Gibbs, R.A.
TITLE Direct Submission
REFERENCE Submitted (22-NOV-1996) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza S930, Houston, TX 77030, USA
FEATURES
SOURCE Location/Qualifiers
1. 1198
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="23920, 23921"
/sex="female"
/tissue_type="brain"
/dev_stage="infant"
/clone_lib="Scares library INIB from IMAGE consortium"

BASE COUNT 380 a 216 c 250 g 352 t
ORIGIN

Query Match 21.0%; Score 326.4; DB 11; Length 1198;
Best Local Similarity 99.4%; Pred. NO. 4.4e-65;
Matches 338; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
```

Db 1152 CCAGTACCAGATGCCAACAATCATGATTGCTGACCTGATGAGACAGAGTGAAATAAAT 1093  
Qy 1272 gaattacaagaagattattacattcatctggtttagacttaattgcacaacgaccac 1331  
Db 1092 GATTTACAAAGATATTACATCATCTGGTTTACACTTAATATGACACAGACACAC 1033  
Qy 1332 gaccttcccaaggtgacacccgcctcaagcctgcaagtgagggtgcttcctcaatcaacgagc 1391  
Db 1032 GACCTTCCAGAGTGACACCGCCCTCAGCTCAGAGTGGGGCTGCTCCTCATCAGCGGGGC 973  
Qy 1392 gctgtcccgacgacgacgtcggtcggtgagctgagctgactctgactgagcaga-ctcct 1450  
Db 972 GCTGTCCCGACGACGACGCGGCTGAGCTGAGCTGACTGACTGAGCAGAGCTTCT 913  
Qy 1451 ggtgtatgttccagaagtgtgaaattatgtgttaaatctgactcattcgtatgcta 1510  
Db 912 GGTGTATTTTTCAGAAATGCTTGAATGTATGTGTTTAAATCTGCTCATTCGATGCTA 853  
Qy 1511 ggtatacatatgatttcaataaatagaacttttaaga 1550  
Db 852 GGTATACATATGATTTTCATTAATATGAATCTTTTAAAGA 813

RESULT 4  
G27428 576 bp DNA STS 28-JUN-1996  
LOCUS human STS SHGC-11921, sequence tagged site.  
DEFINITION G27428  
ACCESSION G27428  
VERSION G27428.1 GI:1396147  
KEYWORDS STS; STS sequence; primer: sequence tagged site.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Myers,R.M.  
1 (bases 1 to 576)  
Unpublished (1996)

REFERENCE  
AUTHORS  
JOURNAL  
COMMENT  
Contact: Richard M. Myers  
Stanford Human Genome Center (SHGC)  
Stanford University School of Medicine  
Department of Genetics, M-344, Stanford, CA 94305, USA  
Tel: 4157259687  
Fax: 4157259689  
Email: myers@shgc.stanford.edu

Primer A: TATTTTCCAGTATACCCAGCA  
Primer B: TCGTTGTGCAATTTAAGTC  
STS size: 127  
PCR Profile:  
Initial incubation: 94 degrees C for 90 seconds

Denaturation: 94 degrees C for 15 seconds  
Annealing: 62 degrees C for 23 seconds  
Polymerization: 72 degrees C for 30 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9600  
Protocol:  
Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
Tag Polymerase: 0.05 units/ul  
Total Vol: 10 ul

Buffer:  
MgCl2: 2.5 mM  
KCl: 50 mM  
Tris-HCl: 20 mM  
pH: 8.3

Prepared with primer pairs provided by Sandoz, derived from R54092  
-- Washington University/Merck EST sequence.  
Location/Qualifiers

source 1. .576  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
STS  
primer\_bind 24. .150  
primer\_bind 24. .45  
BASE COUNT 144 a 123 c 111 g 190 t 8 others  
ORIGIN

Query Match 16.6%; Score 257.2; DB 13; Length 576;  
Best Local Similarity 95.1%; Pred. No. 3.8e-49;  
Matches 330; Conservative 0; Mismatches 9; Indels 8; Gaps 6;

Qy 1212 ccagttcccaagat-gccaaatctatgtctgaact-ggatgaaacagatgaataa 1269  
Db 32 CCAGTTACCCGACATNCTTCAACAATCTGATGCTGACCTGGATGGAACAGAGTAAATAA 91  
Qy 1270 atgaattacaagaagatattacattcatct-ggtttagacttaataatgcacaacgac 1328  
Db 92 ATGATTTACAAAGATATTACATTCATCTGGGTTTACACTTAATATGCCACAGCCAC 151  
Qy 1329 caagacattccca-gggtgaacccgcctcagcctgcagtgagtggtgctcctcaacgac 1387  
Db 152 CACGACCTTCCAGAGGGTGACACCGCCTCAGCCTGAGTGGGCTGAGCTCATCAACGC 211  
Qy 1388 gggcgctgtcccgacgacgagtgagg--ctgagctgagtgctgactctgagcaga 1445  
Db 212 GGGCGCTGTCCCGACGACGAGTGGGCTGGGAGCTGAGCTGAGCTGAGTGAAGAGA 271  
Qy 1446 -ctccgtgtatgatttcaagaatgctgaagtatgttttaactctgctcattcg 1503  
Db 272 GCTCTGGGTGTATGTTTTCAGAAATGCTTGAAGTATGTATGTTTAAATCTGCCATTCG 331  
Qy 1504 tatgtatgattacatatgatttcaataaatagaacttttaaga 1550  
Db 332 TATGTAGTATATATGATGATGATTTTCAATAATGAATCTTTTAAAGA 378

RESULT 5  
166494/c 166494 7218 bp DNA PAT 28-DEC-1997  
LOCUS Sequence 14 from patent US 5670367.  
DEFINITION 166494  
ACCESSION 166494  
VERSION 166494.1 GI:2724471  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 7218)  
AUTHORS Dornier,F., Scheiflinger,F. and Falkner,F.Gunter.  
TITLE Recombinant fowlpox virus.  
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;  
FEATURES  
source 1. .7218  
location/Qualifiers

BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others  
ORIGIN

Query Match 5.8%; Score 90.6; DB 5; Length 7218;  
Best Local Similarity 6.9%; Pred. No. 1.1e-10;  
Matches 30; Conservative 252; Mismatches 151; Indels 0; Gaps 0;

Qy 421 tctgacagttgaggaagagctgagagagcgagcgagcgaacttgaagaagaacttga 480  
Db 1496 TGATGCGATCATCTGATATATCTATCTCAAGTACTTAAGAGATGAAGAAATTTGCT 1437  
Qy 481 tctgacagagaaagagggccattgagaagaacatgattgaaaaagaaataagaagaag 540  
Db 1436 ACRRR 1377

FEATURES

Oy		541 agggagtaacatggtataccaagaagtgtgatcttgcagaatatgcccaactygaggcc	600
Dy		..... : : :	
Dd		1376 RR	1316
Oy		601 caggtaggaagaagttacaagaagattcagcgcatctaataoacttgaggaaatcaa	660
Dd		1316 RR	1257
Oy		661 agccagctgcttcgcgggaatgacctcacaaaagtgtgtggaagaatcgctcatacg	720
Dd		1256 RR	1197
Oy		721 ctgataaacacaacatgysgaagatgtagcgcsaaaaaggacagacagatgcagaca	780
Dd		1196 RR	1137
Oy		781 aaactaacaggagctcttgtaatataagatcaggaatatgagaattggaatagaccg	840
Dd		1136 RR	1077
Oy		841 gatgaagaaccaac	853
Dd		1076 RRRRRRRRRRRRRATC	1064
RESULT	6		
ID	E08995	standard; RNA; UNC; 1686 BP.	
XX	E08995;		
SX	E08995.1		
DT	07-OCT-1997 (Rel. 52, Created)		
DT	07-OCT-1997 (Rel. 52, Last updated, Version 1)		
Xx		DNA encoding a protein involved in Leucocytozoon phytlaxis.	
KW	JP 1995089995-A/1.		
OS	unidentified		
OC	unclassified.		
[1]			
RN	1-1686		
RP	Kato A., Onaga H., Ueda S.:		
RA	"POLYPEPTIDE INDUCING IMMUNITY AGAINST LEUCCYTOZOON PROTOZOOAN AND RECOMBINANT DNA MOLECULE CODING THE POLYPEPTIDE";		
RT	Patient number JP 1995089995-A/1, 04-APR-1995.		
RL	DOUBUSUYOU SEIBUTSUGAKUREKI SEIZAI KYOKAI, NITSUSEIKEN KK.		
XL			
OS	Leuccycytozoan caulleryi		
CN	PN JP 1995089995-A/1		
CC	PD 04-APR-1995		
CC	PI 10-SEP-1993 JP 1993226078		
CC	PF KATO ATSUSHI, ONAGA HIROSHIE, UEDA SUSUMU		
CC	PC C07K14/44,A6LK39/015,C12M1/21,C12M15/09,C12P21/02,G01N33/53,		
CC	G01N33/569,		
CC	(C12M1/21,C12R1:19),(C12P21/02,C12R1:19);		
CC	strandness: Double;		
CC	topology: Linear;		
Key	Location/Qualifiers		
FT	source	1..1686	
FT	/organism="Leuccycytozoan caulleryi"		
FT	/note="mRNA is derived from 2nd shizont"		
FT	mat_peptide	1..1686	
FT	/product="a protein involved in Leuccycytozoon phytlaxis"		
TH	key	Location/Qualifiers	
TH	source	1..1686	

FT	/db_xref="taxon:32644"
FF	/organism="unidentified"
xx	
sq	Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T; 0 other;
	Query Match      4.3%; Score 67.2; DB 24; Length 1686; Best Local Similarity    45.1%; Pred. No. 2,5e-05; Matches 249; Conservative 0; Mismatches 303; Indels 0; Gaps 0;
OY	328 cagcaagaagaaagcgaagtgcctatgtaacaagtgaaacaagtttgcataatatctgaagaa 387
Dd	71 CAATTAATAAATAATTAAAAATGTGAAAAAGAAAGAAGAAAATGAAGAAATATGAAAAAGAG 130
OY	388 gccaatcttgaaaaaaccaaggcctttaaccagtgtagccaggtltgaaggaagacgtggag 447
Dd	131 AACCAAGAACGAC 190
OY	448 aggcagcgcygagcgaacttgaaaagaacttcgatccagcaagaagaaagggccaattgag 507
Dd	191 AACCAAGATGAGAGAAC 250
OY	508 aaagacatgatagaaagaaagaaataacgaagaagaagagagtcatayggaatcaaagatgtg 567
Dd	251 AAGAAAGAAAGAAC 310
OY	568 atcttgtctccgaatattgtgcccaacttgtagagcccagcggtggaagaaaggttaacaagaagaaag 627
Dd	311 AAGAAAGAAC 370
OY	628 attcaagcatattaatcaactggagggaagaaattcaaaagccagcttgccttcggaagaaatggat 687
Dd	371 AAGAAAGAAC 430
OY	688 gtcacaagaagtggtgtgaggaagaaatgcgcgtatacagcttgaaataaaccacacatygaagagat 747
Dd	431 AAAATGAAGAAC 490
OY	748 gagsgaagaagaagagcacagagaggttcgagcgcaaaacctaacagaggtcttggaaatraaa 807
Dd	491 AAGATGAC 550
OY	808 gatcagaagaatagagaaattgagaatagaactgtagtgaagaaacaaacacacttgyaagac 867
Dd	551 AACCAAGAAAGACTATATCTCGTGAAGAAAGAAATGAAGATGAGAGAGAGAGAGAGAGAG 610
OY	868 gaggcagcagaag 879
Dd	611 AAGAACACAGAGAG 622
RESULT	7
LOCUS	AC007461/c DNA PRI 04-JUN-1999
DEFINITION	Homo sapiens chromosome 17, clone 84_E_24, complete sequence.
ACCESSION	AC007461
VERSION	AC007461.8 GI:4996920
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. Britten,B., Linton,L., Nussbaum,C. and Lander,E. 1 (bases 1 to 180385) Castele,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cook,P., DeRellano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gaede,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., Baker,J., Baldwin,J., Barna,N., Beckery,R., Benn,J., Brown,A., Calkins,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A., Cook,P., DeRellano,K., Depayre,E., Devon,K., Dewar,K., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gaede,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,

Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L., Karatas, A., Lehoczy, J., Lieu, C., Locke, K., MacDonald, P., McGowan, N., McGowan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Nilotf, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

Direct Submission  
Submitted (04-MAY-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 180385)

REFERENCE  
Birken, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, D., Barna, N., Beckerly, R., Benn, J., Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearlano, K., Depayre, E., Devon, K., Dewar, K., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L., Karatas, A., Lehoczy, J., Lieu, C., Locke, K., MacDonald, P., Marquis, N., McGowan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Nilotf, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

TITLE  
JOURNAL  
Submitted (04-JUN-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 4, 1999 this sequence version replaced gi:4966400.  
All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

Only 180385 base pairs from the middle of this clone are being submitted. The remainder overlaps either accession number AC005152 (WICGR project L335) or accession number AC005243 (WICGR project L343).

FEATURES  
source  
location/Qualifiers  
1. .180385  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="84.E.24"  
/clone\_1lb="Alan Buckler -- per comm"  
/map="17"  
/chromosome="17"  
446. .753  
/rpt\_family="AlusC"  
2375. .2419  
/rpt\_family="(TTTGTG)n"  
complement(2985. .3103)  
/rpt\_family="L2"  
complement(3146. .3352)  
/rpt\_family="L2"  
complement(3531. .3639)  
/rpt\_family="L2"  
complement(3798. .4394)  
/rpt\_family="L2"  
5531. .5631  
/rpt\_family="MER81"  
5648. .5674  
/rpt\_family="(CAAAA)n"  
5812. .6020  
/rpt\_family="L1MC4"  
6180. .6224  
/rpt\_family="AT\_rich"  
6303. .6611  
/rpt\_family="AlusP"  
7378. .7618  
/rpt\_family="MIR"  
9593. .9623

/rpt\_family="(GGCA)n"  
complement(10666. .10923)  
/rpt\_family="AlusG"  
10958. .11088  
/rpt\_family="MIR"  
11752. .11914  
/rpt\_family="L2"  
complement(11915. .12219)  
/rpt\_family="AlusX"  
12220. .12347  
/rpt\_family="L2"  
13110. .13182  
/rpt\_family="GA-rich"  
13526. .13564  
/rpt\_family="(TCC)n"  
13575. .13660  
/rpt\_family="(TA)n"  
complement(14380. .14658)  
/rpt\_family="AluO"  
14742. .14830  
/rpt\_family="(TA)n"  
14860. .14967  
/rpt\_family="(TA)n"  
complement(14977. .15122)  
/rpt\_family="L2"  
complement(15292. .15385)  
/rpt\_family="L2"  
16672. .16781  
/rpt\_family="L2"  
16802. .16846  
/rpt\_family="(TGA)n"  
17225. .17245  
/rpt\_family="(A)n"  
18411. .18720  
/rpt\_family="AluB"  
complement(20715. .20788)  
/rpt\_family="MLT1J2"  
complement(20822. .20916)  
/rpt\_family="MLT1J2"  
20956. .21137  
/rpt\_family="MIR"  
21830. .21872  
/rpt\_family="AT\_rich"  
21957. .22257  
/rpt\_family="AlusX"  
complement(22438. .22538)  
/rpt\_family="MLT1J"  
complement(23375. .23667)  
/rpt\_family="AluY"  
24553. .24858  
/rpt\_family="AluY"  
complement(25021. .25234)  
/rpt\_family="MIR"  
26957. .27261  
/rpt\_family="AluY"  
complement(27524. .27669)  
/rpt\_family="MLT1B"  
complement(27670. .27964)  
/rpt\_family="AlusP"  
complement(27965. .28210)  
/rpt\_family="MLT1B"  
complement(28228. .28522)  
/rpt\_family="AlusG"  
complement(28592. .28700)  
/rpt\_family="MIR"  
complement(29353. .29537)  
/rpt\_family="MIR"  
29802. .30095  
/rpt\_family="AlusG"  
complement(30099. .30219)  
/rpt\_family="MIR"  
30227. .30267  
/rpt\_family="(TG)n"

```

repeat_region      30276..30321
                    /rpl_family="(CA)n"
repeat_region      complement(30322..30611)
                    /rpl_family="AluDo"
repeat_region      complement(30956..31151)
                    /rpl_family="MER3"
repeat_region      complement(31533..31633)
                    /rpl_family="L2"
repeat_region      31810..31858
                    /rpl_family="L2"
repeat_region      32028..32107
                    /rpl_family="L1P4"
repeat_region      complement(32919..33182)
                    /rpl_family="MIR"
repeat_region      34169..34314
                    /rpl_family="CT-rich"
repeat_region      34774..35007
                    /rpl_family="MIR"
repeat_region      35589..35621
                    /rpl_family="AT-rich"
repeat_region      35763..36142
                    /rpl_family="MER41A"
repeat_region      36143..36452
                    /rpl_family="AluSg"
repeat_region      36453..36636
                    /rpl_family="MER41A"
repeat_region      38011..38071
                    /rpl_family="CT-rich"
repeat_region      38083..38262
                    /rpl_family="(TTC)n"
repeat_region      38235..38319
                    /rpl_family="(TTC)n"
repeat_region      38320..38499
                    /rpl_family="(TTC)n"

```

Query Match 4.2% Score 65.8; DB 40; Length 180385;  
 Best Local Similarity 45.2% Pred. No. 6.2e-05;  
 Matches 241; Conservative 0; Mismatches 292; Indels 0; Gaps 0;

```

OY 347 cttatgaacaggtgaacaagtttgcataatcttgaggaagccaatttgaaaaaacca 406
DB 38635 CATATTCAAACTGAGATCTGCGAGAAATAGAAAATAAATAAGAGAAGAAGAACA 38576
OY 407 aggccttaatccaagtgtgacccggttgaggaaggaagtcgagagcgagcgagcacttg 466
DB 38575 AGGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38516
OY 467 aaaaagaacttcacatcagcaagaagaagggccattgagaagacatgatgaaaaag 526
DB 38515 AGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38456
OY 527 aaatacgaagaagaagagatcatgtggaatcaagaatgttgatctgtctcagaatat 586
DB 38455 AGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38396
OY 587 cccaacttgagagcccaagtggaagaaggttacaaaggaagaagattccaactataatca 646
DB 38395 AGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38336
OY 647 tggaggaatccaagccagcgtcttcggaagaatgatgtccaagaagtggttgag 706
DB 38335 AGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38276
OY 707 aaatgcgtatcagctgataaaccacaatgagagaagatgagcagaagaagaagcaca 766
DB 38275 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38216
OY 767 gagatccaaggaacaaactaacacaggtatcttgaatataaatcaaggaataaggaat 826
DB 38215 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 38156
OY 827 tgagatatgaactgtagaagaagaacaacacttgaaacagagcagcagaag 879

```

DB 38155 AGGAG 38103

RESULT 8

CELC3368 39103 bp DNA INV 08-JUL-1998

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 39103)

Wilson, R., Alnsough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Cope, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Showkeen, R., Smalton, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Wellstock, L., Wilkinson-Sproat, J., and Wohlman, P.

2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

Nature 368 (6466), 32-38 (1994)

94150718

2 (bases 1 to 39103)

Gattung, S. and Wu, X.

The sequence of C. elegans cosmid C3368

Unpublished (1998)

3 (bases 1 to 39103)

Waterston, R.

Direct Submission

Submitted (29-MAR-1996) Robert Waterston

4 (bases 1 to 39103)

Waterston, R.

Direct Submission

Submitted (08-JUL-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

Submitted by:

Genome Sequencing Center

Department of Genetics, Washington University,

St. Louis, MO 63110, USA, and

Sanger Centre, Hinxton Hall

Cambridge CB10 1RQ, England

e-mail: twenematode.wustl.edu and jesesanger.ac.uk

COMMENT

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

#### NEIGHBORING COSMID INFORMATION:

The 5' cosmid is T07H8, 200 bp overlap; 3' cosmid is F20A1, 2000 bp overlap. Actual start of this cosmid is at base position 1 of CELC3368; actual end is at 6614 of CELF20A1

#### NOTES:





COMMENT	Draft entry and computer-readable sequence for [1] kindly submitted by R.T.Ogata, 12-SEP-1989.
FEATURES	..location/qualifiers
SOURCE	1. .17522



[illegible][illegible]







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 00:31:03 ; Search time 6491.04 Seconds

(without alignments)  
-432.366 Million cell updates/sec

Title: US-09-502-945-2  
Perfect score: 2885  
Sequence: 1 gggaatcccttctgtcgaagt.....aaacaaaagttaaaattt 2885

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 segs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pl2:\*  
11: gb\_pl3:\*  
12: gb\_ro:\*  
13: gb\_ro:\*  
14: gb\_sy:\*  
15: gb\_sy:\*  
16: gb\_vt:\*  
17: gb\_vt:\*  
18: em\_fun:\*  
19: em\_hum1:\*  
20: em\_hum2:\*  
21: em\_cm:\*  
22: em\_cm:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vt:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pr5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2885	100.0	2885	40 AF039691	AF039691 Homo sapi
2	2762	95.7	3369	40 AF132608	AF132608 Homo sapi
3	2375.4	82.3	3450	12 AF207748	AF207748 Mus muscu
4	2375.4	82.3	3531	12 AF006602	AF006602 Mus muscu
5	2181.6	75.6	2233	9 AB011172	AB011172 Homo sapi
6	953.6	33.1	3255	40 AF132607	AF132607 Homo sapi
7	953.6	33.1	8459	9 AB006626	AB006626 Homo sapi
8	668	23.2	3758	11 HSM800965	AL117455 Homo sapi
9	625.8	21.7	2871	12 AF207749	AF207749 Mus muscu
10	289	10.0	120766	40 AC004150	AC004150 Homo sapi
11	222.4	7.7	4238	9 AB018287	AB018287 Homo sapi
12	217.2	7.5	55893	56 AC023855	AC023855 Homo sapi
13	214.6	7.4	54589	42 AC014460	AC014460 Drosophila
14	214.6	7.4	182015	55 AC010993	AC010993 Drosophila
15	203.8	7.1	1385	40 AF124924	AF124924 Homo sapi
16	173.6	6.0	3648	40 AF132609	AF132609 Homo sapi
17	173.6	6.0	4078	9 AB020708	AB020708 Homo sapi
18	173.6	6.0	4099	9 HSA011972	AL137696 Homo sapi
19	173.6	6.0	4854	11 HSM802148	AF137696 Homo sapi
20	163	5.6	3488	35 AF128433	AF128433 Drosophila
21	163	5.6	3843	12 AF006603	AF006603 Mus muscu
22	150.4	5.2	4435	4 XLMILZ	Z97214 Xenopus lae
23	135.2	4.7	32480	34 CELC10E2	AF026202 Caenorhab
24	134.6	4.7	14737	43 AC017824	AC017824 Drosophila
25	134.6	4.7	113060	43 AC008255	AC008255 Drosophila
26	111.6	3.9	2750	8 AF064207	AF064207 Schizosac
27	111.6	3.9	40438	7 SP041410	U41410 Schizosacch
28	90.8	3.1	451	12 MOSIR3E2	M10668 Mouse DNA w
29	90.4	3.1	38686	34 CEFA3G6	Z50070 Caenorhabdi
30	90.4	3.1	52689	32 CECS4H10	Z81488 Caenorhabdi
31	90.4	3.1	153835	32 CEY6D1	AL023673 Caenorhab
32	90.4	3.1	159933	55 AC022705	AC022705 Homo sapi
33	89	3.1	55893	56 AC023855	AC023855 Homo sapi
34	88	3.1	122186	11 AC004466	AC004466 Homo sapi
35	87.6	3.0	159933	55 AC022705	AC022705 Homo sapi
36	87.4	3.0	100923	51 AC002124	AC002124 Human BAC
37	87.2	3.0	2723	7 SCYNT021W	Z71297 S.cerevisia
38	86.2	3.0	97348	12 AF091216	AF091216 Mus muscu
39	84.6	2.9	3489	16 K5U52064	U52064 Kapost's sa
40	84.6	2.9	32207	5 AR065852	AR065852 Sequence
41	84.6	2.9	137508	16 K5U76698	U75698 Kapost's sa
42	83.4	2.9	49745	43 AC017028	AC017028 Homo sapi
43	83	2.9	5719	12 MNCARS1	AF051726 Mus muscu
44	82.8	2.9	133661	16 U93872	U93872 Kapost's sa
45	82.6	2.9	460	12 MOSIR3E1	M10296 Mouse DNA w

## ALIGNMENTS

RESULT 1  
AF039691 2885 bp mRNA PRI 05-MAY-1999  
LOCUS Homo sapiens antigen NY-CO-9 (NY-CO-9) mRNA, partial cds.  
DEFINITION AF039691  
VERSION AF039691.1 GI:3170181  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 2885)  
AUTHORS Gordon,J.D., Tureci,O., Sahin,U., Pfreundschuh,M. and Old,L.J.  
TITLE Characterization of human colon cancer antigens recognized by  
autoantibodies  
JOURNAL Int. J. Cancer 76 (5), 652-658 (1998)  
MEDLINE 98272252  
REFERENCE 2 (bases 1 to 2885)  
AUTHORS Scanlan,M.J.  
TITLE Direct Submission  
JOURNAL Submitted (23-DEC-1997) Ludwig Institute for Cancer Research, New  
York Branch at Memorial Sloan-Kettering Cancer Center, 1275 York  
Avenue, New York, NY 10021, USA  
FEATURES  
source location/Qualifiers  
1..2885  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="colorectal carcinoma"  
gene <1..2885  
/gene="NY-CO-9"  
<1..2695  
/gene="NY-CO-9"  
/note="putative Type II integral membrane protein"  
/codon\_start=2  
/product="antigen NY-CO-9"  
/protein\_id="AAC18040.1"  
/db\_xref="GI:3170182"  
/translation="EFLSKSEPTPGSLNHSILPQHPKCGAHNSILDOSSPGSP  
GTPSYRLPDPGYDSKDDPLKRTASPEMLKRSRKQVVAERSSPLRRKDYI  
STRKRAVEITGAPGASVCNSAPGSPSPNSHSTIAENGESVPVPIEMLPQ  
HRAIPDSSPNQSLYTPSLIPNLSGLQATVYTNHSLAEPILQOEAERQALOS  
LRGGTLTGKFMSTSIIPGCLLQVLAEGDSPGHASILQHVLLLEQARQOSTLAVP  
LHGOSPLVTEGERTVSMRTVQKLPBHRNLRTOSSPLPOSPALQOLVMOOHOFLF  
KORQOLQGLKILTKGELPPOPTTPEHEPEELTBOEVLLEGALTMRESSES  
STQDLEEDBEDEBEREDCTQVKDEEGSGAEBCDLEPGAGTKLPSDAQPIOP  
LQYQVAPLSLATVPHQALGRTOSSPAAPGSKNPDPVKNHFTTSVYDTFMLKQC  
MGNTNHPHAGRIOSIMSRLOETGLSKCERIKGRKATLDEIQTVHSEYHTLKQT  
SPLNROKLDSKLLGIPSOKMAYALPCGGIGVSDVYWNMHSASAVRAVAGCLIEA  
FKVAAGELKNGFALIRPGHAEESTAMGCFENSVATKLLQOKLVNGVAILYMD  
IHNGNGQAFYNDPSVLYLSLRHSDYNGPSSGAEFGVGVYNNVNTAGV  
DPRIGDEVITARTVYMPYIAHESRPVYVSAFGDAVBHLSPLGISTYAKCFGL  
TROLMTLAGRVVLAEGHDLTAICDASACVALLSVLPQIDEVLAQKPNINAV  
ATLEKYEIISKMSVCQKFAAGISLRGAQAGEETEEAM"

BASE COUNT 626 a 901 c 838 g 520 t  
ORIGIN

Query Match 100.0%; Score 2885; DB 40; Length 2885;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaattctctgtcgaaagtaagagagccacacccagcgagcctcaacattcctctcc 60  
|||||  
Db 1 GGAATTCCTCTTTCGAGTCAAAGAGACCCACACAGCGGCGCTCAACCATTCCTCTCC 60  
QY 61 acaagacacacaaatgctgagagacccacatgctctcttgaccagagattccctccca 120  
|||||  
Db 61 ACAGCACCCCAAATGCTGGGAGGCCACCATGCTTCTTGACACAGAGATTCCCTCCCA 120  
QY 121 gagcgcccccttggagagcctcctctacaacatgctcttgcctgggacctacgacag 180  
|||||

Db 121 GAGGGCCCCCTGGGAGCGCTCCCTCTACAACTGCTTTGGCTGGGCCCTACAGACAG 180  
QY 181 tcgaagcagacttcccccctccgcaaaacagcctctgaaccacattgaagtgcgttaag 240  
|||||  
Db 181 TCGAGACACACTTCCCCCTCCGCAAAACAGCCTCTGAACCCACTTGAAAGTGCCTTCAAG 240  
QY 241 gctaaaacagaaggttgctgtagcaggaagcagtcctccctctgctgcgcaagatgtagac 300  
|||||  
Db 241 GCTAAACAGAAAGGTGGGTGAGCGGAGAGAGATGCCCTCTCGGTGGCCAAAGATGGGAC 300  
QY 301 tgtattatgaacacattaaagaagaagctgtttagatcaacaggtgcgaggtgcgcgc 360  
|||||  
Db 301 TGTATTATGACACTTTAAGAAGAGAGCTGTGTAGATCTACAGATGCCGGCTGGGGGCTTC 360  
QY 361 gtccgtgtagtaacagagcagccggtctccgcccagcctcccaacagctcccaagac 420  
|||||  
Db 361 GTCCGTGTAAACAGCGACCGCGCTCGGGCCCAAGCTCTCCAAACACTCCACAGAC 420  
QY 421 catcgctagaatgagcttacttgctcagatccccaacatacccaactgagatgctccctca 480  
|||||  
Db 421 CATGCGTGAAGATGGCTTTACTGGCTCAGTCCCAACATCCCACTGATGCTCCCTCA 480  
QY 481 gaaccagacccctccctctgagacagctccccaacacagttcaagctctcaacgctctctc 540  
|||||  
Db 481 GCACCGAGCCCTCCCTCTGGACAGCTCCCAACACAGTTACAGCCCTCTACAGCTCTCTTC 540  
QY 541 tctgcccacaactccctcctcagagctgcaagagcaggtcaactctcaacactcaacactca 600  
|||||  
Db 541 TCTGCCCAACACTTCTCCTTAGGGCTGCAAGGCAAGGTCACTTCAACACTACACTCTAC 600  
QY 601 tgcctcccccgaagctgtgcaacacagcagagagcaggaagcagccctccagctctgcg 660  
|||||  
Db 601 TGCCCTCCCGAAGCTGTCGACACAGAGAGAGCCGAGAGGAGGAGGAGGAGGAGGAGGAGG 660  
QY 661 gcaaggtgagcagctgtagccgagcaagttcatgaagacatctctattctctgctcgtcgt 720  
|||||  
Db 661 GCAGGGTGGCAGCTGACCGCGCAAGTTCTATGAGCAACCTCTATCTCTGCTCTCTCT 720  
QY 721 ggaggtgagcactgtagagagcagcagagccccaagagatgacccctcgtcgtcagagatgt 780  
|||||  
Db 721 GGAGGTGGCACTTGAAGAGGCGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 780  
QY 781 gctgttctgtagaagagcccgagcagcagagacacctcatltgctgtgcaactccacgagca 840  
|||||  
Db 781 GCTGTGTTGGAGAGAGGCGCGGAGAGAGACACCTGCTGTCGACATCCACGGGAGCA 840  
QY 841 gtccccactagtagcaggtgtgaagctgtgtgccacagaatcgcgcagctgaagcaagctccc 900  
|||||  
Db 841 GTCCCCACTAGTAGACGGGTGAACGTGTGGCCACAGCATGCGGAGGAGGAGGAGGAGGAGG 900  
QY 901 ggcgcacatgcgcccctgagcgcgacatgaatctctcaaccgctgtagcagagattcccaagccct 960  
|||||  
Db 901 GCGGATCGGCGCCCTGAGCCGCACTCACTGCTTCAACGCTGCGGAGAGATCCCAAGGCCCT 960  
QY 961 gcaagcagctgtgtcatgcaacaacagcagcagcagltccctgtagaagcagcagcagca 1020  
|||||  
Db 961 GCAGCAGCTGCTCATGCAACAAAGACACACAGCATCTCTGAGAAGAGAGAGAGAGAGAGCA 1020  
QY 1021 gctacagctgtagcaagatctcaccagaagaagagagctgcccagcagagccaccaccca 1080  
|||||  
Db 1021 GCTACAGCTGGGCAAGATCTCCACCAAGACAGGGAGAGCTGGCCAGGACCCACACCCA 1080  
QY 1081 cccctgagagacagaagagagctgtagcggagacagagaaggtcttctgtgtaggagagagc 1140  
|||||  
Db 1081 CCTTGAGAGACAGAGAGAGAGCTGACGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
QY 1141 cctgacacatgcccgggagagagctccacagagagtgtagcacaaggaaggaagccttgagga 1200  
|||||  
Db 1141 CCTGACACATGCGCCGGGAGGAGCTCCACAGAGAGTAGAGAGACACAGAGAGAGAGAGAG 1200  
QY 1201 ggaggaagaggaagagatgtaggagagagagagagatgtcatccaggttaagagcagga 1260  
|||||  
Db 1201 GGAGGAGAGAGAGAGAGATGGGGAGAGAGAGAGAGATTCATCCAGGTAAAGAGAGAGAG 1260

OY 1261 gggcgagagtggtgtcgaagagggcccgacttggaggaagccttggtgtcgtgatacaaaaa 1320  
|||||  
DB 1261 gggcgagagtggtgtcgaagagggcccgacttggaggaagccttggtgtcgtgatacaaaaa 1320  
OY 1321 actgttctcaagatgcaccaacgcgttgaaccttgcagtgtaaccaagcgcctcaagcct 1380  
|||||  
DB 1321 ACTGTTCTCAGATGCGCAACCGCTGGAACCTTGTGACGGTGTACAAAGCCGCCCTCAGCCT 1380  
OY 1381 ggcacatgtgcccaacaaagccctggcgctaaccaatcctccctgtcgtcccttgggg 1440  
|||||  
DB 1381 GGCCTACTGTGCCCAACCAAGCCCTGGCGGTACCCATCTCCCTGCTGCCCTGGGGG 1440  
OY 1441 catgagaaccccccaagacaaacccgttaagacacctcttcacaaagaagtgtgtacaga 1500  
|||||  
DB 1441 CATGAGAACCCCCCAAGACCAACCGTCAAGCACTCTTACACACAAAGTGTGTACGA 1500  
OY 1501 cagcttcatgtcaaaagcaacagtgcatgtgcgggaacacacacgttcacccctgaagcatgc 1560  
|||||  
DB 1501 CACGTTTCATGCTAAAGACACCAAGTCATGTGGGGAACACACAGTGCACCCGTGAGCATGC 1560  
OY 1561 tggccgagatccagagatcgtgtcccgctgcagagagacagcctgtctagaagtgcga 1620  
|||||  
DB 1561 TGGCCGAGTCCAGACATCTGCTCCGCGCTGCAGGAGACAGGCTGTGAGCAAGTGCGA 1620  
OY 1621 gcgagatccgaagtgccgaagccagcgtatgatatccagacagtgcatctcgaatacca 1680  
|||||  
DB 1621 GCGAGTCCGAGGTGCCAAAGCCACGCTAGATGATCCAGACAGTGCACCTGTGAATACCA 1680  
OY 1681 caccctgtcctatgggagcagtcctcctcaacgcggcagaagctcagaacagaagaattgtct 1740  
|||||  
DB 1681 CACCTTGCTCTATGGAGACAGTCCCTCTCAACCGGACAGAAAGCTAGACAGCAAGAAGTGTCT 1740  
OY 1741 cgggtcccatcagccsagaatgtatgtgtcgtcctgtgtgggggcatcgggggtggacag 1800  
|||||  
DB 1741 CGGTCCCATCACCCAGAGAGATGATGCTGTGCTGCTGTGGGGGATGGGGGTGGACAG 1800  
OY 1801 tgaacccgtgtgagatgagatgactcctcagtgctgtgscgcaatggcagtggtgcgtcct 1860  
|||||  
DB 1801 TGACACCGTGTGAATGAGATGACTCTCCAGTGTGTGCCATGTGCGTGGGCTGCT 1860  
OY 1861 gctgagagctgagctcctcaagtggtcgtcgaagagagctcacaagatgtaattgcataccg 1920  
|||||  
DB 1861 GGTGAGGCTGGCTTCAAGGTGGCTGCAGAGAGAGCTCAAGATGATTTGGCCATCATCCG 1920  
OY 1921 gcccccagagacacaaagccgaggaatccacagccatggaattcgtcttcaactcgt 1980  
|||||  
DB 1921 GCCCCAGAGACACACAGCGCGAGGAATCCACAGCATGGGATTTCTGCTTCAACTCTGT 1980  
OY 1981 agccatcaccgaaaaactcctcagcagaagaagtgaacgtgggcaagtgctcaatcgtga 2040  
|||||  
DB 1981 AGCCATTCACCCGAAAACCTCTACACAGAAAGTTGAACGTGGCAAGGTCTCATCGTGA 2040  
OY 2041 cgggagacatcaacatggaatgagcaccagcagcgttctcaatgagccctcgtgtgt 2100  
|||||  
DB 2041 CTGGGACATTCACCATGGCAATGCAACCCAGCAGCGCTTCTCAATGACCCCTCTGTGCT 2100  
OY 2101 ctacatctctcgtcactgcgtatgacaacggaactcttccaggcctcgtgggctcctga 2160  
|||||  
DB 2101 CTACATCTCTCTGCATCGCTATGACACGGAACCTTCTTCCAGGCTCTGGGGCTCCTGA 2160  
OY 2161 agaggttggtgagagacagcgtgtgggtgataatgtgaacgtgagatgagagaggtgt 2220  
|||||  
DB 2161 AGAGGTTGTTGAGAGACACAGCGGTGGGTACATGTAACGTGGCAAGGTGAAGGTGT 2220  
OY 2221 ggaaccccccatgtgagacgtgtgagtaaccttaccagcctcagaagcgtgtgagccat 2280  
|||||  
DB 2221 GGAACCCCCCATGTGAGAGAGCTGAGTACCTTACAGCTTCAAGACAGTGTGATGCCCT 2280  
OY 2281 tggcacaagatctcaactgattgtccttagtctcccgaggtttgattgtctgttgaag 2340  
|||||  
DB 2281 TGCCACAGAGTTCTCACCTGATGTGTCTACTCTCCGCGGTTTGATGTGTGAAGG 2340

OY 2341 acatgtctcctctgtggtgactctgtcaaccgacagatgttggccacttgaacag 2400  
|||||  
DB 2341 ACATGTCTCTCTGTGGGTGGCTACTGTGTCAACGCGCAATGTTTGGCCACTTGAACCA 2400  
OY 2401 gcaagtgaaccccttgcagagggccgggttggtgtcgtcccttgaagggagcattact 2460  
|||||  
DB 2401 GCAGGTGATGACCCCTGCGAGGGGGCCGGGTGTGTGCGCTGAGAGGAGGCATGACTT 2460  
OY 2461 gacgcacatctgtatgtcctcgaagctgtgtctcgtcgtcgtcgtatgtaagtga 2520  
|||||  
DB 2461 GACCGCATCTGTGATGAGCTCTGAACTGTGTCTGTGCTGTCTGCTACGTAAAGCTGCA 2520  
OY 2521 gcccttggatggagcagttcttcagcaaaaagcccaatcaacagatggccaagctaga 2580  
|||||  
DB 2521 GCCCTTGATGAGGAGTCTTGACGCAAAAGCCCAATCAACGAGTGGCCAGCTAGA 2580  
OY 2581 gaaagtcaatcagatccagaacaaacacgtgagctgtgtgcagaagtctgcgcgtgtgtc 2640  
|||||  
DB 2581 GAAAGTCATCGAGATCCAGGCAACACCTGAGCTGTGTGCAAGATTTGCCGCTGTCT 2640  
OY 2641 gggcgggtccctgcagagggcccaagcagtgagaccgaagaagccgaaatgtgaagcc 2700  
|||||  
DB 2641 GGGCGGTCCCTGCGAGGGGGCCCAACAGTGAACGGAAGCCGAATGTGAACGCC 2700  
OY 2701 atggccttctgtgtgtgtggggccggaacagcccaagctgtgcgagccgggaacagc 2760  
|||||  
DB 2701 ATGGCTTCTGTGTGTGGGGGCCCAACAGGCCCAAGCTGTGCGCACCCGGGAACACAC 2760  
OY 2761 cccagcgcgagagagccatgagcagagagcctgtccctgtgaagcccgcccat 2820  
|||||  
DB 2761 CCCAGSGCGGAGAGAGGCCCATGAGAGAGAGAGCTGCTGTGAGGCCCGGCCCAT 2820  
OY 2821 cccttggcctcaccatgtgtattgtttatcttctatataaacaagaagttaa 2880  
|||||  
DB 2821 CCTTGGGCTTCCACCATGTGATTTGTATTGTTTCTATTAATAAACAAGATTAA 2880  
OY 2881 aattt 2885  
|||||  
DB 2881 AATT 2885  
RESULT 2  
AF132608  
LOCUS AF132608 3369 bp mRNA PRI 06-MAY-1999  
DEFINITION Homo sapiens histone deacetylase 5 mRNA, complete cds.  
ACCESSION AF132608  
VERSION AF132608.1 GI:4754908  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 3369)  
Grozinger C M., Hassig C.A. and Schreiber S.L.  
Three proteins define a class of human histone deacetylases related  
to yeast Hda1p  
Proc. Natl. Acad. Sci. U.S.A. 96 (9), 4868-4873 (1999)  
JOURNAL  
MEDLINE  
99238449  
2 (bases 1 to 3369)  
Grozinger C M., Hassig C.A. and Schreiber S.L.  
TITLE  
Direct Submission  
Submitted (03-MAR-1999) Chemistry and Chemical Biology, Harvard  
University, 12 Oxford Street, Cambridge, MA 02138, USA  
JOURNAL  
FEATURES  
location/Qualifiers  
1..3369  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Jurkat T cell"  
1..3369  
/note="HDAC5; similar to Saccharomyces cerevisiae Hda1p"  
/codon\_start=1  
/product="histone deacetylase 5"  
/protein\_id="AAD29047.1"

/db\_xref="GI:4754909"  
/translation="MNSPNESDGMGSRPSLEILPRTSLHSIPVTEVKPLPRAMPS  
SMGGGSPSPVELRGALVSDPTREOOLLOELALKKOOLLOLOLEAEOKOH  
DHLRHOEYLOLKH.KOOEMLAKROOELROORROORROORROORROORROOR  
EOULLIRNKEKSKESIASIEYLRLOEPLSLSKREPTREGLNLSLPOHKKGAH  
ASLDQSSPPOSPPPTPKLPLGPPDSKDDPLKRTASPEMLKRSRKOAVAR  
RSSPLLRKDGTVISTEKKRAVEITGAGSVSNAGSPSSPSHSTIANEPK  
TGSVPNIPTMLPQHRALPDLSSPNOSLYSPSLNPSISLQATVTVNHLASPK  
LSTQOAEAROLALROGGLTGKFNSTSPGCLLVLESDSPGHSLSLOHLL  
LEORROOSTLIAPLHSGPLTGERVATSRMTYKLPRIHPLSRLSPPOSROL  
OOLYMOOHOOLEKOKOOLKILFKGELPROPTHEPEEPELETRQVLLGE  
GALTRPREGSTESTOBDLEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE  
GKRLFSDAOPLQVQAPLSLATVPHOALSTGSPAPGSKSPDQPVNHLPT  
TGVVYDFMLKHQCMGNTHYHPRHAGRIOSIMSLQETGLGICSEIRIGRKATLDEI  
QTVSEYHTLLYGTSPINROKLDKILGIPSGMYAVLPDGGIGVSDPYVMNHHSS  
SAVMAVGLLELAFKVAAGELKNGFALIRPGHHAPESTMGCFPNVSVALITKLLD  
OKLNVGVLIVDMIDHNGNQCAFYNDPSVLYLSLRINDGNFPERSGAREEGSP  
GVGINVAVMTGVPPIGDEYELTARIVYMPYIAHESPDVYLVASGFDVBEHLS  
LGYSVYARCFGLRQLMTLAGRVVLAEGHDILAIDASACYSALLSVELQPL  
DEALVQKPNINAVATLEKYLEIOSKHMCSYOKFNAOIGRSLREAOGETEAEATVSA  
MALTSVGEAOAAAREHSRPRAEFMEQEPAL"

BASE COUNT 727 a 1038 c 1039 g 565 t  
ORIGIN

Query Match 95.7%; Score 2762; DB 40; Length 3369;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 2785; Conservative 0; Mismatches 20; Indels 1; Gaps 1;

OY 1 ggaatctcttgcgaagtaagaagaccacacagcgagcctcaaccattccctcc 60  
DB 564 GGAATTCCTCTGTCGAAGTCAAGAGACCCACACAGCGGCGCCCAACATTCCTCTCC 623  
OY 61 aacagcaccacaatgctggagagccaccatgctcttggacagaattccctcccca 120  
DB 624 ACAGACCCCAAGTGGGAGGAGCCACCATGCTCTTGAGACAGATTCCTCCGCCCA 683  
OY 121 gagcgcccccttggagagcctccctctacaactgcttgcctgggacctaaagac 180  
DB 684 GAGGCGCCCGCTGGAGGCGCTCCCTCTACAACTGCTTGCGCTGGGCGCTACAGAC 743  
OY 181 tcgagaagaactcccccctcgcaaaacagccttgaaccaacttgaagtgcgttaag 240  
DB 744 TCGAGACACTCTCCCGCTCCGCAAAACAGCCTTGAAACCTTGAAAGTGCCTTCAAG 803  
OY 241 gctaaacagaagtggttgagcgagaagcagtcacctctgctgctgcaagatggac 300  
DB 804 GCTAAACAGAGGAGGCTGAGCGGAGAGCAATCCCTCTGCTGCAAGATGGGAC 863  
OY 301 tgtatttagcactttaagaagaagctgttgaatcacaagtgccggagcctggggcg 360  
DB 864 TGTATTATGACACTTTAAGAGAGAGCTGTGAGATCAAGATGCGGCGCTGGGCGCTC 923  
OY 361 gtcgctgtgttaacagcgacccggcctcgccccaagctctcccaacagctcccaagac 420  
DB 924 GTCCGTGTGTAAACAGCGACCCGGCTCGGCGCCAGCTCTCCCAACAGCTCCACAGAC 983  
OY 421 catcgctgagaatgcttactgctcagctcccaaatccccaactgagatgctcccca 480  
DB 984 CATGCGTGAAGATGGCTTACTGCTCAGTCCCAACATCCCACTGAGATGCTCCCTCA 1043  
OY 481 gcaacgagccctccctctggaacagctccccaacagcttcaagctctcaagctctcc 540  
DB 1044 GCACCGAGCCCTCCCTCTGAGACAGCTCCCAACAGATTTAGCCCTTACAGTCTCTTC 1103  
OY 541 tctgtcccaacatctccctaggtgtgcaagcgacaggttcaactgtcaacaaactcaac 600  
DB 1104 TCTGCTCCCAACATCTCCCTAGGCGTGCAGAGCCACGCTCATCTCAACAACTCAAC 1163  
OY 601 tgcctcccggaagctgttgacacagcgagagcggaagagcgagcctccagctccctgg 660  
DB 1164 TGCCTCCCGGAAGCTGTGACACAGCAAGAGGCGGAGAGGCGCTTCACGCTCCCTCG 1223

OY 661 gcaaggctgcaagctgacccggacaggttcaatgagcacatctctatctctgctgctgct 720  
DB 1224 GCAAGGTGGACAGCTGACCGGCAAGTTCATGAGACATCTCTATTCCTGGCTGGCTCT 1283  
OY 721 gggcgctgcaatgagagggcgagcgggagcccccaggggacatgctctctctgctgcaag 780  
DB 1284 GGGCGTGGCACTGGAGGGCGAGGGAGGCGCCACCGGATCCCTCCGCTGCAAGCAAGT 1343  
OY 781 gctgtgttgtagcaaggcgccggcaagcagaccctcatctgtctgcaaccctcaaggac 840  
DB 1344 GCTGTGCTGAGACAGAGCGCCCGGACAGACAGCACTCTCATTTGCTGTGCTGCTCA 1403  
OY 841 gtccccaatgtagcaggggtggaacgttggccaacagcatctggaacggttaggcaagctcc 900  
DB 1404 GTCCCACTAGTAGAGGGGTGAACGTTGTGGCAACAGCATGCGGAGAGGTAAGCAAGTCC 1463  
OY 901 gggcgatggccccctgagcggcagctcagctctcaacgctgctgcaagagatccccaagg 960  
DB 1464 GCGGATGGCGCCCTGAGCCCGCACTCATCTCACCCGCTGCGGAGAGATCCCGAGGCCCT 1523  
OY 961 gcaacagctgtgcatgcaacaacagcaccagcagcttccctgtagaagaagcaagcaagca 1020  
DB 1524 GCACGAGCTGTCTATGCAACACAGACACACAGCATTTCTTGAGAGAGCAAGAGCAGCA 1583  
OY 1021 gctacagctggcgaagatctctcacaagaagcagggagctgcccagcgagccccacacca 1080  
DB 1584 GCTACAGCTGGGCAAGATCTCTACCAACAGACAGGGGAGCTGGCCAGGCGCCACACCA 1643  
OY 1081 cccctggaagagaagaagagagcttgaagagagcagcaaggaagcttctgtctggggagggag 1140  
DB 1644 CCGTGAAGAGACAGAGGAGGAGCTGACGAGACAGCAGAGAGGCTTGTGCTGGGGAGGAGC 1703  
OY 1141 cctgacacatgccccggggagggctccacagagagtgagagcacaaggaagacccctggagga 1200  
DB 1704 CCTGACATGCTGCCCGGGAGGGCTCCACAGAGATGAGAGCACAAGAGACCTGGAGGA 1763  
OY 1201 ggaaggaaggaagaagagatgaggagaggaagagatgtcaatccaaagttaagaacgagga 1260  
DB 1764 GGAAGAGCAGAGAGAGATGGGAGAGAGAGAGGATTCATCCAGGTAAAGAGAGAGA 1823  
OY 1261 gggcgagaagtgctgtcgtgagagggggccgacttggaggaagcctgtgtgatacaaaaa 1320  
DB 1824 GGGCGAGAGTGTGCTGAGGAGGGGCGCGAGCTTGAGAGAGCTGTGCTGATCAAAA 1883  
OY 1321 actgttccagatgcccaacggctgcaaccttgcaggtgtgacaagagggcccccaagcct 1380  
DB 1884 ACTGTCTCAATGCTCCAGCGGCTGTCAGCTTTGTGAGGTGTAACAGCGGCGCTCAGGCT 1943  
OY 1381 ggcacatgtgcccccaacgaagccctgggacgctgacccaatctccctctgtgccccctgggg 1440  
DB 1944 GGCACATGTGCCCAACAGCGGCTGGGCGGTAAACCAAGTCTCCCTGCTGGCGGG 2003  
OY 1441 catgaagaaccccccaagaccaaaccgctcaagcacttcaacaaagtgtgtgtctaga 1500  
DB 2004 CATGAAGAGCCCCCAGACAGCGGCGGTCAACACCTCTTCAACAGAGTGTGCTCA 2063  
OY 1501 caagcttaatgtaagaacaggtgcatgtgtgggaacacaacagtgcaacccctgagatgc 1560  
DB 2064 CACGTTATGCTTAAGACACAGCTGATGTGGGAGAACACACAGCTGCACTTGACATGC 2123  
OY 1561 tggcggatcacaagcatctgtgccggctgcaaggaagagcctgtcttaagaatgca 1620  
DB 2124 TGGCGGATCCAGAGCATCTGTGCTCCGGCTCAAGAGACAGGCTGTGTAAGAGTGGGA 2183  
OY 1621 ggggataccgaagctgcgaagcgaagctagatgagatccagaacagtgcaactgtgatacca 1680  
DB 2184 GCGGATCCGAGAGTGCAGAAAGCACAGTAGATGAGATCCAGACATGTGCACTGTGAATCA 2243  
OY 1681 caacctctctatggaagcagctccctcaacggcgaagaagctagaacagaaagagctgtct 1740  
DB 2244 CACCTGCTCTATGGAGCACTCTCCCTCAACCGGAGAGAGCTTAACAGCAAGAGATGCT 2303  
OY 1741 cgtgccatcagccagaagatgtatgtctgtcctgtgtggggcatcggggtggaag 1800



Db	2304	CGGCCCCATACGCGAGAGATGTATCTGTGCTGTGCTGTGTGGGGGCATCGGGGTGGACG	2363
Qy	1801	tgacaccgtctggaatbaga tgcacccctccag tgc tgcgcacggaagtgggctgcct	1860
Db	2364	TGACACGCTGTGAATTAAGATGCACTCTCDOACTGTGTGGCAGATGGCAGTGGCTGCT	2423
Qy	1861	gctggaagctgagcctccaag tggctcgcaagagaagctcaagaatlgatltgcatcatcgc	1920
Db	2424	GCTGGAGCTGGCCTTTCAAAGTGGCTGACAGAGAGCTCAGAAATGGATTTGGCATTCATCG	2483
Qy	1921	gcccccgagacacacgcgcgaagaatccaaagccatgggaattctgcctctccaactcgt	1980
Db	2484	GCCCCCAGGAAACCCGCCAGGAATTCACAGCCAGGAGATTCCTCTTCATCCTGT	2543
Qy	1981	agccatcacgcgaanaactctcatagacagaagt tgaacg tgggcgaagfcccacatcgt tga	2040
Db	2544	AGCCATCACCCGCAAAACTCTTACAGCAAGAA GTTGAACGTGGCAGAGTCTCATCTGTGA	2603
Qy	2041	ctggagacatboaaccat tggcaat tggca cccgaagcgcgtctctcaat tga cccctc tgc t	2100
Db	2604	CTGGAGCATTTACCATATGGCAATGGACCCAGCAGCGGCTCTATATATGACCCCTGTGCT	2663
Qy	2101	ctaatctctctgcattgcattgcata tga caa cgggaactcttccaagctctgggctctga	2160
Db	2664	CTAATCTCTCTGCATGCTATGCTATGACACGGAACCTTCTTCCAGGCTCTGGGGCTCTGA	2723
Qy	2161	agaagtttgg tggagaggaacgaagcgt tggggtgacaa tctgaaacgt tggat tga cgaaggggt	2220
Db	2724	AGAGCTTTGTTGGAGGACCAAGCCTGGGGTTCATATGTAACGTGGCATGTGACAGAGGTGT	2783
Qy	2221	ggaaccccccat tggagaag tggagfacct tcaagccttcaagaacagt tggat tgc ccat	2280
Db	2784	GGACCCCCCATTTGGAAACGTGAGTACCTTACAGCCTTCAAGACAA GTGTGATGGCCAT	2843
Qy	2281	tggccacgaagttctcaactga tgt tgg tcc tagtctccgcggg tttg at b cgt tga ag	2340
Db	2844	TGCCACGAGATTCTCACTGATGTGGTCCAGTCTCCGCGGGTTGATGCTGTGAAGG	2903
Qy	2341	acatcttctccctctgg tgg tgc taccctct taa ccgcgaatgttttggccacttggacag	2400
Db	2904	ACATCTGTCTCTCTGGTGGTGACTCTGTCAACCGCAATATTTTGGCACTTGACCG	2963
Qy	2401	gcaagctatgaaccttgcgaaggggcgggt tgg tgc tgc gcc tggaa gggagccatgaact	2460
Db	2964	GCACCTCATACCCCTGGCAGGGGGCCGGGTGTGCTGGCCCTGGAGGGAGGCCATGACTT	3023
Qy	2461	gaaccgcacatctgtatgtcct tgaagct tgtgtctcgcgtctgcag tgtaaagct tga	2520
Db	3024	GACCGCATGTGTGATGCTCTGAGGCTTGTGTCGTGGCTGTGCTCACTGATGAGACTCA	3083
Qy	2521	ggccttggat tgaagcagct tct tgcga caa aagcccaacat caa cgcga g tggccaogctga	2580
Db	3084	GCCCTTGATGAGGCACTCTTGACGAAAAAGCCCAATCAACGAGTGGCCAGCCTAGA	3143
Qy	2581	gaaagcatcagagatccaagcgaacaac tggagcgt tgc tgcagaag t tccgcgc tgc t	2640
Db	3144	GAAAGTCAATGGAATCCAGAGCAAAACTGTGAGCTGTGTGCAAGAGTTTCCCGCTGGTCT	3203
Qy	2641	gggcgcgttccctgcgaagggcccaacgaag tgaagccgaagaagccga -aat tgt tgaagc	2699
Db	3204	GGGCGGCTCCCTGCGAAGAGGCCCAACAGAGTGTGAGACCGAGGAGCCGAGACTGTAGAGGC	3263
Qy	2700	catgagcttgc tgt tgg tgggggcga aagcccaagctgcgcgaagcccggaacacag	2759
Db	3264	CATGGCTTGTCTGTGCTGGTGGGGCCAGAGAGGCCAGCTGTGGCGAGCCCGGAACACAG	3323
Qy	2760	cccccaagccgcgaagagagccca tggagcaggaagcct tgcct tga	2805
Db	3324	CCCCAGGCGGCGAGAGAGCCCATGTGACGACGAGAGCTTGCCTGTGA	3369

[illegible]

OY	121	gagcgcccccttggcgagcgctccctccctctactcaaacgctcttgcttggcgccttaagaag	180
Db	660	GAGGGCCCTCTCTGGAGCGCTCCCTCTCTCAAAATTGGCTTTGGGGCCCTATGAAAG	719
OY	181	tcgagacgacttccccctccgcgaacagcctcttgaaccaacttgaagctgcttcaag	240
Db	720	CCGTGATGACTTTTCCCTCCGTAAACGGCGCTGGAAACCCTTAAGTACGTTCCAG	779
OY	241	gctaaacaagaagctggcttgaagcagatccctctcgtcgcgaagatlyggac	300
Db	780	GCTTAAACAGAAAGGTATCCAGAGAGAAACATCCCTCTCGCGCAAGAAAGATGGCAC	839
OY	301	tgattatagccctttaagaagaagcgtgttgaatcacaaagtgcggcgccctgggggctt	360
Db	840	TGTTATTATGACTTTTAAAGAAAGAGACAGTTGGATCACCGGCAGGGGCTCGGGGTGTC	899
OY	361	gtccgctgtltaacagcgccaccggtctccgcccacgtctcccaacagctcccaagaac	420
Db	900	GTCGGTGTGTAAACAGTACGGCCGGCGCTCGGCCACCACTCTCCAAACAGTTCCACAGCAC	959
OY	421	catgcgttgaagatgcttctactgtgtctcaagtcaccaatccccacatgagatgtctccca	480
Db	960	CATGCTGAAGACGGCTTTACTGTGCTCAAGTCCCAACATCCCACTGAGATGATATCCCA	1019
OY	481	gcaccagccccctccctcttgcagcagctcccccacaacagttcaagccttacaagctctcttc	540
Db	1020	GCACCGGGCCCTCTCTGTGGACAGTTCCCAACAGATTCAAGCTCTATAGTGTCTCTTC	1079
OY	541	tctgtcccaacatctccttaagtgcttgcagcgccaagctcaatcttcaaccaactacactac	600
Db	1080	TCTGTCCCAAACTCTCCCTATGGGCTGACAGGCACTGTCACTGTACCAATCCCACTCAC	1139
OY	601	tgcctcccccaagatgtgcagacagcagagggcgccgagagcagccctcaagctccctcg	660
Db	1140	CGCTCCCCAAGCTGTCAACACAGCAGAGGCTGAGAGGACGGCCCTTCAGTCCCTCG	1199
OY	661	gcagaggttgcagcgttgcacggcgcaagtcatlyagacacatcctatctcgtgctgcgt	720
Db	1200	GCAGGGCGGCGACACTGCACGGCAAGTTCAATGACACATCTCAATCCCTGCGTGCCTGT	1259
OY	721	ggcggttgcacttgaaggcgagcgaggccccacgggcatctgctccctgctgagcagatgt	780
Db	1260	GGGAGGTGCACTGTGAGGCTATCACAAAGCCCCACGGGCAAGCTTCCCTGTGTCACACAGT	1319
OY	781	gctgttgcgtgagagaagcccgcgacagagacacctcatctgtgttgcacctccaagggca	840
Db	1320	TTTGCTCTCGGACA---GGCCGGCACAAAGACACAGCTCATAGCAGTGGCCGCTTCATGGGCA	1376
OY	841	gtcccaactagtgaacggttgaacgcttggccaccacagcatgaggaacgtgacgaagctccc	900
Db	1377	GTCCCCACTGTGTGACGGGCTTAAGGTGTGGCCACACATGAGAGAGCGTGGTAACTCTCC	1436
OY	901	gcgcgactcgcccccttgcgcgacactcagctccctcaacgcgtgcgcgacaggtccccaagccct	960
Db	1437	GAGGACACGAACCTCTGAGCGCGCACTAGTCTCTCCCGCTGCGGACAGAGTCCCAAGCCCT	1496
OY	961	gcacgagctgtcatctgacacaacagcaccaagcagatctcttgaagaaagcaagcaagca	1020
Db	1497	GCACGAGCTGTGATCATGCAGCAGCAGCACACAGCAAGTCTCTGGAAGAACAGAAAGCACACA	1556
OY	1021	gtctacagcttggcaaggtctctacccaagaagggagctgtgcccaagcagccccaacacca	1080
Db	1557	GATCAAGCTGGGCAAGATCTTTACCAAAATGGGAGAGCTGTCAAGGCAAGCCCCACACATCA	1616
OY	1081	cccttgaagagacagaaggaagccttgcacggagcagcagagagcttctgtcggggagggagc	1140
Db	1617	CCCGAGAGAGACAGAAAGAGAGCTGACGGAGCAGCAGCAGAGGCTTGTCTGGAGAGGGGCGC	1676
OY	1141	ccctgacacatgccccggagagtgctccacagagatgtgagagacacagaagaagacttgaaga	1200
Db	1677	CCCTAACATTTTCCCCGGGAAAGGCTCTACAGAAAGTGAGACACCCAGAGAAACCTTAGAAGA	1736

[illegible]

Db	2814	GCCTATGCCCCAGGAGTTCTCACTGAGCTGCTCTCACTGCTCGCTGGGTTTGATGCTGT	2873
Oy	2335	tgaaggaacatctgtctccctcttggtgtgctactctgtcaccgcagatgttttgcacatt	2394
Db	2874	TGAAGACATCTGCTCTCCACTGCGTGGGCTATTTCTGTCAACGCCAATATGTTTGGCCACTT	2933
Oy	2395	gaccagcagctgtatgcacccttgcagggggccgggtggtgtctgtcccttggaggagccca	2454
Db	2934	GACCAGCAGCTCATGACACTGCTGCGGGGGCCGGGTGGTGGCTGGGCCCTGGAGAGGCCA	2993
Oy	2455	tgacttgaccgcacactgtgtatgctctcttgaagcttgtgtctctgtctgtcgtatgttaa	2514
Db	2994	TGACTTGAACCCGCACTCTGGAAGCTCTGAGGCCCTGTCTGTGCTGTGCTTGCCTGACGCTGA	3053
Oy	2515	gtgtgagcccttgtgagtgagcagctcttgcagcaaaagcccaactcaacgcagttggcac	2574
Db	3054	GCTGCAGCCCTTGGATGTAAGCAGTCTTTCGACAGAAAACCCAGCTCATGCGTTGCCAC	3113
Oy	2575	gctagaagaatgcatcagagatccacaagaacaactgtgagtgtgtgcagaaagttgcgcgc	2634
Db	3114	ACTAGAAGAATGATCATCGAGATTCAGAGCAAAACACTGAGACTGTGTACAGAGTTTGCCGC	3173
Oy	2635	tggtctgggcgggtccctctgcagagggcccaagcaggtgagaccgaagaagccga-aatgt	2693
Db	3174	TGCTGTGGGCTGTCTGCTGCTCGGGAGGCTCAGACAGTGTGAAGAAGAGGAGGCCGAGACTGT	3233
Oy	2694	gaagccatgtgacctgtcgt	2753
Db	3234	GAGCCCATATGCCCCCTGCTTTCCTGTGGGGCTGTGAGAGGCCCAAGGCTGTGGCCACTCAGCA	3293
Oy	2754	acacagccccaagccgcgcagagagagcccatgtgagcagagcctctgccttgcagccgcgc	2813
Db	3294	GCACAGCCCCAGAGCCAGAGAGAGAGCCCATGTGAGCAGAGAGGCTGCTGTCACACCTTG	3353
Oy	2814	cccccatcccttgggtctcaacattgtgatttgtttatttttctatataaacaaca	2873
Db	3354	CCCCCATCCCTTGGGCTTCATCATGTGATTTGTTATTTTCTATTAATAAACAAAA	3413
Oy	2874	agttaaaatt 2884	
Db	3414	AGTCACACATT 3424	
RESULT 4	AF006602		
LOCUS	AF006602	3531 bp	mRNA
DEFINITION	Mus musculus histone deacetylase MHD1 mRNA, complete cds.		15-FEB-2000
ACCESSION	AF006602		
VERSION	AF006602.2	GI:6978313	
KEYWORDS	.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 3531)		
AUTHORS	Verdel,A. and Khochbin,S.		
TITLE	Identification of a new family of higher eukaryotic histone		
	deacetylases. Coordinate expression of differentiation-dependent		
	chromatin modifiers		
JOURNAL	J. Biol. Chem. 274 (4), 2440-2445 (1999)		
MEDLINE	99107904		
REFERENCE	2 (bases 451 to 3531)		
AUTHORS	Verdel,A. and Khochbin,S.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JUN-1997) Institut Albert Bonniot, INSERM U309, La		
	Tronche 38706, France		
	3 (bases 1 to 3531)		
REFERENCE	Verdel,A. and Khochbin,S.		
AUTHORS	Direct Submission		
TITLE	Submitted (15-FEB-2000) Institut Albert Bonniot, INSERM U309, La		
JOURNAL	Tronche 38706, France		
REMARK	Sequence update by submlttr		

COMMENT On Feb 15, 2000 this sequence version replaced gi:415859.  
 FEATURES  
 source  
 1..3531  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /tissue\_type="total fetus"  
 /dev\_stage="embryo 13.5-14.5 dpc"  
 85..3426  
 /note="similar to yeast histone deacetylase HDAC1"  
 /codon\_start=1  
 /product="histone deacetylase mHDA1"  
 /protein\_id="AA09834.2"  
 /db\_xref="gi:6978314"  
 /translation="MNSPESDGMGSRPSLGIIPRTPLHSIPVAEVKPYLPAMPSP  
 SMGGGGGSPVELRGALPMDPLREOLOLOELVLKQOOLLOLLEAFQKOH  
 DHLRHOVLOLKHLLKQOQEMLAARKOLEEOROEOROEORLEKRLLEOLLINLN  
 KEKESKESIASTEYVKRLRDLERFLSKSEPTTGGTNGHNSIPQPKKMGANHSIDSSPP  
 QSGPPTSPKTLPLGPGYDSRDPLFNKTASENLKVRSLKQKVARERSPLLRRK  
 DGYIISTEPKRAVEITGPGVSSVSNAPSDGSSPSSHSITAEANGFTGSVNIPTPT  
 EMIDQHRALPLDSSPNQFSLTSPSLSPNLSIGLQATVYVNHSLTASKISTOGEARR  
 OALOSLRGGFTLGEPMSTSIIPGLGLVALGEPTSPGHAHSLHONHSCVSGROSTT  
 IAVLHGOSPLVTGERVATSMRTGKTLPRHPLSRTPSSLPJOSPOLLOOLYMOOQHD  
 QPLEKQOQOMQLKITLKTLELSKOPTLRPEETBEELTDEQELLGEGALITRBESS  
 TESSSTODELEEEREEEREEEDCIVKDEEGEGPDEGPOLEESSAGKYLLEADQO  
 LQPLQVYAPLSLATVPHQALGRTPSOAETSPQVTVVKKHLFTTGTVVDTFPM  
 LKQMGCGNTVHPHEHAGRIQSIINRSLOETFLSKCEIPDQTKVHLITVHSEYHRT  
 CEGHRLRLMLLAGGRVVALLEGHDLTALIDABECVSAALLVELQPLDEAVLQQR  
 SVNVAITLEKYEITQSKHMSCVQRFPAAGLIGSLREAOITGEKEEFVTSAMALLSVGAEE  
 QAAVAVTDEHSPRAPEEMQEPPL"  
 BASE COUNT 792 a 1071 c 1038 g 630 t  
 ORIGIN  
 Query Match 82.3%; Score 2375.4; DB 12; Length 3531.  
 Best Local Similarity 89.8%; Pred. No. 0;  
 Matches 2597; Conservative 0; Mismatches 281; Indels 13; Gaps 4;  
 Db 1 ggaattcctctgtgcgaagtcgaagagagcccaacacagcgagcctcaaccattccctccc 60  
 621 ggaatttcctgtgtgcgaagtcgaagagagcccaacacagcccgagcctcaaccattccctccc 680  
 QY 61 acagcaacccccaaatgctctggggagagcccaacatgctctcttttggaacagaagttccctccc 120  
 Db 681 acagcaacccccaaatgctctggggagagcccaacatgctctcttttggaacagaagttccctccc 140  
 QY 121 gggggggcccccttggaacgctcctcctcaaaaatgctctgtgcttgagcccttaagaag 180  
 Db 741 gggggggcccccttggaacgctcctcctcctcaaaaatgctctgtgcttgagcccttaagaag 200  
 QY 181 tggagagagcattcccccttcgcgaacaaagcctctgaaccacattggaagtgctgttc 240  
 Db 801 cgggtatgactttcccccctcgttaaaagccctcgggaacccacttaaaagtacgttcgag 260  
 QY 241 gctaaacagaagatggtctgtagcggaagaagaagtcctccctcctgctgcgcgaagatggagc 300  
 Db 861 gctttaaagcagaagtgtagcggaagaagaagtcctccctcctgctgcgcgaagatggagc 320  
 QY 301 tgttatgaaccttaagaagaagagcgtttgagatcaagaagtgcccgagccctggaggtc 360  
 Db 921 tttttattgacttttttaagaagaagacacttttaagatcacccggacggggccttggggttc 380  
 QY 361 gtcctgtgttaacagcgcaacccggctccggccccagctctcccaacagctcccaacagc 420  
 Db 981 gttccgtgttaacagcgcccgccgctcggcccaacgctctcccaacagttcccaacagc 1040  
 QY 421 catgcctgagaatgagcttaactgtgctcaagtcceccaacatcccactgagaatgctccctca 480  
 Db 1041 catgcctgagaatgagcttaactgtgctcaagtcceccaacatcccactgagaatgctccctca 1100

QY	481	gcaccagaccctccctctgtagcagctcccccaccagcttcagctctacagctctcctc	540
Db	1101	GCACGGGGCCCTCCCTCTGGACATTTCCCAACCAAGTTACGCTTTAACTCTCTTC	1160
QY	541	ctctgcacaacatccctctaaagagctgcagagcagagctcagctgttacaactacactac	600
Db	1161	TCTGCCAACAATCTTCCTTAGGGCTGCAGGGCACTGTCTACTGTCCACCACTCCCACTCAC	1220
QY	601	tgctcccccagaagctgttcagacaacagcagagcgagagcgagccctcagctccgcg	660
Db	1221	CGCCCTCCCGAAGCTGTCAACACAGCGAGGCGCTGAGAGGGAGGCCCTTCAATCCCTCGCG	1280
QY	661	gcagaggtgcacagctgcagccggaagtctcatgagcacatctctatctccggtgctctc	720
Db	1261	GCAGGGCGGCACTTGACGGCAATTTCAAGACAAATCTTCATCTTCATCTCGGTGCTGTT	1340
QY	721	gggcgtgcacatgcagagggcgagcgagccccccagggcagctccctctctgcacagct	780
Db	1341	GGGAGTGGCACTTGGAGGGGTGACAAAGCCCCCAGGGGGCAGGTTCCCTGTGTCACACAGT	1400
QY	781	gctgtgtctgcagcagggcccgagcagagacccctcatctgtctgtgcactccagggca	840
Db	1401	TTTGCTCTGGACA---GGCGGGCAACAGACACAGCTCATATAGAGTGGCCCTTCAAGGGCA	1457
QY	841	gtccccaactagtcagcgggtgaacgctgtgcagccacagcatgcagctgcagctcc	900
Db	1458	GTCCCCACTGGTGAAGGGGTACAGCTGTGGCCACACACATGAGAGACGGTGGTMACTCC	1517
QY	901	gcgcagctgcgccctgcagcgcgacatcagctccatccgcgtgcgcagagatcccaagccct	960
Db	1518	GAGGACAGCACTCTGTGACCGGCACATCAAGTCTCTCCGCTGGCGAGAGTCCCAAGGCCCT	1577
QY	961	gcacagagctgtctcagaaaacagcagcagagctctcccgagagacagaaacagcacca	1020
Db	1578	GCAGCAGCTGGTCAATGCAGCAGCAGCACAGCAAGTTCTGTGGAGAACAAACACACACACA	1637
QY	1021	gtctacagctgcggcaagaatctctccacaagaagaggagctctgcagcgagcccaacaccca	1080
Db	1638	GATGCAGCTGGGCAAGATCTTTACCAAAATCGGGGAGCTGTCAAGGACGCCACACATCA	1697
QY	1081	ccctgcagagagacagagagagagctgcagcagcagcagcagagagctctgctgcgggagagagc	1140
Db	1698	CCCCGAGAGACAAACAAAGAGAGCTGACGGAGCGAGGAGGCGCTTGTGGGAGGGGCG	1757
QY	1141	ccctacacatgcgccggagagggctccacagagaaatgcagacacacagagaaagactgcagga	1200
Db	1758	CCTGACATTTCCCGGGAAAGGCTTACAGAAATGAGAGCACCCAGAAAGACCTAGAA	1817
QY	1201	ggaagagcagagagaagagatgcagcagagagagagatgtcatccaggttaagcagagga	1260
Db	1818	GGAGAGAGAGAGAGAGAGA---GGAGGMAAGGAGGAGCTCATTTCAAGGTCAAGAGAGAGA	1874
QY	1261	ggggagagagtgctgcagagagagggcgccgaactgcagagagagccttgctgcagataaaaa	1320
Db	1875	TGGCGAGAGTGTCTCTATTAAGGCCCTGTCACTTAGAAGAGTCACTGCTGCTTACAAAAA	1934
QY	1321	actgttctcagatgcaccaacgcgtctgaaccttgcagcaggtgtaccagcgccctcaagct	1380
Db	1935	GTTTGTGGCAATATCCCAAGAGTTAAAGCCCTCCAGGTGTACAGGCAACCCCTCAGCCT	1994
QY	1381	ggcacactgtgcaccaaacagccctgcggcgcttaaccgaatctctccctgtctgcgccggggg	1440
Db	1995	GGCCACTGTGCTCATATAGGCGCTTGGGCGCACCCAGTCTCACTCGTGTCTCTGGGAG	2054
QY	1441	catgaagaaccccccaagaacaaacc-----gtcaagacacctcttcacacaagaatggt	1494
Db	2055	CATGAAGAGGCCCAACAAGACCAACCAACTGTGGTGAAGACACGCTTTCACACAGGTTGCT	2114
QY	1495	ctacagacagttcatctctaagacacagatgcatagtctgcgggaaacacacacgctgcacctga	1554
Db	2115	CTATGACACGTTATGTGTGAAGCACCAAGTATATGTCCGGAACACACACAGTGCACCCGA	2174

QY	1555	gcatctggtccggatccaagacatctgtccggctgcagagagacggctctcaagaa	1614
Db	2175	GCACCCGCGCGCATCCAGACATCTTGTCCTCCGACGTGCAGAAACGTGCTCGGCAA	2234
QY	1615	gtcgagacggaatccgaagctgcagagccacgcctagatgatccaaacgctgactctga	1674
Db	2235	GTGTACGGGGATCCGGGGTCTTAAGCCACACTGGATGAATTCAGACCGTGCACCTGCA	2294
QY	1675	ataccacacccctgctctatctggagacagctccctccaaaccgycagaagctagacaagaa	1734
Db	2295	GTACCAACACCTGCTCTATGTGGACACACCCCTTAACCCGACAGAACTGCAGACAGAAA	2354
QY	1735	gttgcttggtcccatcaagccaagaatgtagtctgtctgctgctctgtggggacatcggggt	1794
Db	2355	GCTGTGTGGCCCATTCACAGCCGAAAGATGTACGGCATATCTGCCTGTGGGGGATTTGGGT	2414
QY	1795	ggaacagtacaacccgttgaaatgaaatgaaatctctccagtgctgtgcgaatgagatggg	1855
Db	2415	GCACGTACACGGTGTGGATGTAGATGCACCTCTTAAGTGCCTGTGCATATGGCAGTGGG	2474
QY	1855	cgcgcctctggagcttgccctccaaggtgcgcctgcagagagactccaagaatgaaatgtgcat	1914
Db	2475	CTGGCTGGTGGAGCGCGGCTTCAAGGTGGCTGCAGAGAGAGCTCAAGAAATGGATTGGCAT	2534
QY	1915	catccgcccccaaggaaaccaacgcagagaaatccaaagccaatggatctctctctcaa	1974
Db	2535	CATCCGGCCCCCAGAGACACCTGTGTGAGGTGCACAGCCATGGATTCTGTCTTCTCAA	2594
QY	1975	ctcttgacatctacgcgaaacacctccacagcagaagttgaaacgttggagaaagttccat	2034
Db	2595	CTCCGTACCATTCACAGTAAACTCTCGACAGCAAGCTATGCGGTGGCAAGCTCTCAT	2654
QY	2035	cgctgacgaggaacatccaactgagaaatggacacacagcagcgctctctaaatgacccctc	2094
Db	2655	CGTGACGTGGGATATTTCACCATGGSCAACGGCACCCAGCAAGATTCTTAACAGATCCCTC	2714
QY	2095	tgctctacatctctctgcatcgctatgacaaacggaaattcttccagctctgggc	2154
Db	2715	TGTGCTTACATCTCCCTGCATCCGTTCAGCAACGGGAATCTTTCAGGCTCTGGGGC	2774
QY	2155	tcttgaagaagttgtgtggagagacagcgctgggtatcaatgtgaacgttgcaatgaaacgg	2214
Db	2775	TCTCTAAGAGGTGTGTGGAGGGCCAGGTGTACAAAGTAAAGTGGCTGTGACAGG	2834
QY	2215	agggttgaaaccccccatctgaaagcgtggagataccttaacagctccaagacagtgtagt	2274
Db	2835	AGGTGTGATCTCCCCCATTTGGAGATGTGGAAATACCTGCAGACGCTTCAGACAGTGTGAT	2894
QY	2275	gcccatgtgccacagatctcaaccgtatgtgtgtcctagtctccgcgcgggttbtgatgtgt	2334
Db	2895	GCCCATGTCCCAAGAGTGTACACTGACGTGACGTGCTAGTCTCGCGGTGGTGTATGTCTGT	2954
QY	2335	tgaagaaatctgtctcctctgtgtgtgtactctgtcaacgcagatgttttggccactt	2394
Db	2955	TGAAGGATCTGTCTCTCACTGGGTGGCTATTGTGCACGGCAGATGTTTGGCCACTT	3014
QY	2395	gacacagacagttagaaccttgagagagggccgggtgtgtctctgagcccttgaaagagagca	2454
Db	3015	GACCAAGCATCTTATGACACTGGCTGGGGGCCGGGTGTGTGTGGCCCTGTGAAGAGGCCA	3074
QY	2455	tgaacttgacccaatctgtatgactctgtaagttgtgtctcgcgtctgtcagtgttaa	2514
Db	3075	TGACTTGAACCCCATCTGTATGTGCTGTAGGGCTGTGTCTGGGCTGTGTACAGGTGGA	3134
QY	2515	gtctgacgccttgaatgaagcagctcttgaaacaaagcccaacatcaacgcagtgccac	2574
Db	3135	GCTGAGCCCTTGTGATTAACACAGCTTGTGCAGCAAAAGCCACAGTCAATGGCGTTGGCAC	3194
QY	2575	gtctagaaagaatctgaagatccaagagcaaaacatgagcttgtgagaaagtctgcgcgc	2634
Db	3195	ACTGAGAAATCTATCGAAGATCCAGAGCAAAACCTGGAAGCTGTGTACAGAGGTTGGCCG	3254
QY	2635	tgtgtctggccggtccctctgagagggcccaagcaggttgagacccgaagaagccga-aatgt	2693

Dd	3255	TGCTGGGCTCCTCGCTGCAGGAGGCTACAGAGGTGAAGAAGAGAGCCGAGACTGT	3314
Oy	2694	gaaccatgagccttgctgtgtgtgggacccgaacaggcccagctgcagcgccggga	2753
Dd	3315	GAGGCCCATGGCCCTTCGTTTGGGGGCTTGAGCAGAGCCCAAGCGCTGTTGCCATCACA	3374
Oy	2754	acacagccccagcgcaagagaagcccatgtgagcagaagacctgcctgtgacgcccg	2813
Dd	3375	GCACAGCCCCAGCGCAGCAGAGACCCTTAGAGCAGAGAGCGCTGCCTGTGCACCTGG	3434
Oy	2814	cccccatccctttaggccttcacacatgtgatgtttattttcttattaaacaaa	2873
Dd	3435	CCCCCATCCCTTTGGGCTTCATCATTTGATTGTTTTTAATTTTCTAATAAAAACAA	3494
Oy	2874	agttaaatt 2884	
Dd	3495	AGTCACACTT 3505	
RESULT	5		
LOCUS	ABO11172	2233 bp mRNA PRI 10-APR-1998	
DEFINITION	Homo sapiens mRNA for KIAA0600 protein, partial cds.		
ACCESSION	ABO11172		
VERSION	ABO11172.1 GI:3043723		
KEYWORDS	KIAA0600 protein.		
SOURCE	Homo sapiens male brain cDNA to mRNA, clone_1lb:pbluescriptII SK plus clone:H60785a.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;		
AUTHORS	Primates; Catarrhini; Hominiidae; Homo.		
TITLE	Ohara,O., Nagase,T. and Ishikawa,K. 1 (bases 1 to 2233)		
JOURNAL	Direct Submission Submitted (13-FEB-1998) to the DDBJ/EMBL/GenBank databases. Osamu Ohara, Kazusa DNA Research Institute, DNA Technology; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914) 2 (sites) Nagase,T., Ishikawa,K., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O. Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro DNA Res. 5 (1), 31-39 (1998) 98290545		
REFERENCE	Nagase,T., Ishikawa,K., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.		
AUTHORS	Prediction of the coding sequences of unidentified human genes. IX.		
TITLE	The complete sequences of 100 new cDNA clones from brain which can		
JOURNAL	code for large proteins in vitro		
MEDLINE	DNA Res. 5 (1), 31-39 (1998)		
FEATURES	98290545		
SOURCE	Location/Qualifiers 1..2233 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="HG0785a" /clone_1lb="pbluescriptII SK plus" /sex="male" /tissue_type="Brain" 1..2153 /gene="KIAA0600" <1..2153 /gene="KIAA0600" /codon_start=3 /product="KIAA0600 protein" /protein_id="BAA2526.1" /db_xref="GI:3043724" translation="LRQSGSTLGRKMSNSIPGLIALVREGDSPHASILQHVLLLEQAROSTLIAVPLRGSGSPLYTEGRAVTSMRTYQKLPRHPRLSTOSSP.LPSQPAIQLVNQDHOQLFKOKQQOLQGLIKLTKEELPQPOTPHRETELTEQEVNLGALTPMRGESSTESDELEDDEEDEDEEEDEECIOVKDEGSAGAEQDLBERPGEGYKLFSDADPLELOYAOPLSLATVHQALGTSSPAAPGGM SPDPVHLFLFGTVGVDFPKLHKOCGNTHTVHPHAGIOSIMSWLOETGLSCERISGRKTIDEETGYVAISEVFTLLXGSTPLNRQRLDKILGPISCKMYANLPGGIGYSDPYVMNHSSSAVARNAVCLLEIEAKYAAGELKNGFALIRPQHHAESTMAGROFFSVSAITKKLLOKLVNAGKVLIVMDIHHNGNQQAQFYNPVALYLSLHRDYDNCFPPSGAPEVEGGGTVGVNVNANTGVDPPIGDEVLEYLAFTVMPPIAHNESSPDVVVASGFAVEGHLSL		
CDS	gene		

[illegible]

Dh	901	AGCATGCTGGCCGGATCCAGAGCATCTGTGTCGCCGCTGCAGAGAGACAGGCCCTGTTAGA	960
Qy	1614	agtcgagcggatcccgagtcgcgaagccacgcgtatgatalcagacagtgcactcy	16737
Dh	961	AGTGGACCGGATCCGAGGTGCGAAGCCACCGCTAGATGAGATCCAGACAGTGCACCTCG	10204
Qy	1674	aatacacacccctgctatbgygacagctccctcaacccgacgaagctaaacacaga	17333
Dh	1021	AATACCACACCCCTCTATGTGGACCGTCCCTCAACCGCGAGAACTGACACACAGA	10804
Qy	1734	agttctcgtgcctcatcagcacaagaatgatatctgtgcaccttggggatccggg	17933
Dh	1081	AGTTCTGGGCCCATCAGCCAGAAAGATGATGCTGTGCTTGGGGCATCCGGG	11404
Qy	1794	tgagacagtacacccgtbgyaatalgacatgcctccactcagttgctgtgcatalgacgtg	18533
Dh	1141	TGGACAGTGAACCGGTGTGAGATGAGATGACATCTCCATGCTGTGTCGATGGCAGTGG	12004
Qy	1854	gctgcctcctggaactcggcttcaaggttgctgcagagagatccaaatgtaattccga	19133
Dh	1201	GCTGCTCTCTGAGACTGGCCCTCAAGGTGGCTCAGAGAGCTCAAGAAATGGATTGGCA	12604
Qy	1914	tcattccgcccccaaggacacacagccgaagaatcccaacagcatggatcttcttcca	19737
Dh	1261	TCATCCGGCCCCCAGACACACCGCCGAGAAATCCACAGCATGGATTTCTTCTTCA	13204
Qy	1974	actctgtagccatcaacgcgaanaactccctacagcagaagtltgaactgtggcaagctcca	20333
Dh	1321	ACTCTGTAGCCATCACCAGCAAAATCTCTACAGAGAAATTGAACGTGGGCAAGTCTCA	13804
Qy	2034	tcgtggaactgggaatctcaacaaatgycaaatgycaccagcagtcgtctcaaatgaacct	20933
Dh	1381	TCGTGGACTGGGAACTTACCATGGCATGGACCCAGAGGGGTCTATAATGACCCCT	14404
Qy	2094	ctgtgcctctaatatctctcctcatgcgtatgacaagcgggaactcttccaagctctggg	21533
Dh	1441	CTGTGCTTATATCTCTCTGCACTGCTATGACAGCGGAATCTTTCACAGGCTCTGGGG	15004
Qy	2154	ctctcctaaagagcttggtgtgagagacagcgttgagggtacaaatltgaactgtgcatalgaaag	22133
Dh	1501	CTCCGTAAGAGGTTGTGTGGAGACACAGCGCTGGGGTACAAATGTGAACGTGGCATGGACAG	15604
Qy	2214	gaaggttgagacccccccatctggagaagctggagtaactctacagcctcaagagacgtgtga	22733
Dh	1561	GAGGTGTGGACCCCCCATTTGGAGACGTGGAGTACCTTACAGCTTCACAGACAGTGTGA	16204
Qy	2274	tgcccatggcccaagagtttcaactbgaatgtgttccatgcttccggcggtttgatgtgt	23333
Dh	1621	TGCCATTTGCCACGAGATTTCTCACCTGATGTTGGTCTCATGCTCCGCGGGTTGATGCTG	16804
Qy	2334	ltgaagaagacatctgtccctcctctgggttggtcactctgtcaacgcgcagaatgttttgccact	23933
Dh	1681	TTGAAGGACATCTGTCTCTCTGCTGGTGGTACCTCTTACACCGCAGATTTTGTGGCACT	17404
Qy	2394	tgaccagagcagcttgaatgacccctgagcaagggggccgggtgctgaccttggaaggagcc	24533
Dh	1741	TGACCAGGCAAGCTGATGACCCCTGGCAGGGGGCGGGGTGTGCTGGCCCTCGAGGGAGGCC	18004
Qy	2454	atgacttgacccgcatctgtgatgtgctctcgaagcttggtgtcttcgagctctgctcagtgtaa	25133
Dh	1801	ATGACTTTCACCGCCATCTGTGTGATGCTCTGTGAGGCTTGTGTCTTCGCTCTGCTCACTGTAG	18604
Qy	2514	agctgcagcccttgatgtgatgagcagctcttgagaagaagcccaacaatcaacgaagatgagca	25733
Dh	1861	AGCTGCACGCTTTGGATGAGGCGACTTTGCAGCAAAAGCCCAACATCAACGCAGTGGCCA	19204
Qy	2574	cgtatagaagaatcatcagatcccaagacgaanaaacttgaactgtgtgtgcagaagtctgcg	26333
Dh	1921	CGCTAGAGAAAGTATGATGAGATCCAGAGCAAAACATCTGAGCTGTGTGCAAGATTCGCCG	19804
Qy	2634	ctgtgcttgggccggtctccctcgcagagggcccaagcagatgaagccgaagaagccga-aagt	26923
Dh	1981	CTGTGTGTGGCGCTCTCCCTCGAGAGGCCCAACACAGTATGAGACAGAGAGCGCGAGTCTG	20404

OY	2693	tgaagcgcatacgctctctcttctgtgtgtagggccgaaagccccaagctgtgtagggccgg	2753
Db	2041	TGAGGCGCATGCGCCCTTCTCTCTGCTGAGGGCCGAGCAGGCCCAAGCTGCGGCGACCCGG	2100
OY	2753	aacaacagccccaagccgagcagagagacccaatgaacgaagcctgcctcttaacgcccc	2812
Db	2101	AACACAGCGCCCAAGGCCGCGCAGAGAGGCCCATGAGAGCAGGAGGCTGCGCTGACGCCCG	2160
OY	2813	gcccccacccctttaggcctaccacatgtgatcttcttaattttctatlaaaaaaa	2872
Db	2161	GCCCCATCCCTCTGGGCTCACCAATGTGATTTTGTATTTTTCATTAATAACAA	2220
OY	2873	aagttaaaatc 2884	
Db	2221	AAGTCACACATc 2232	

RESULT	6				
AF132607					
LOCUS	AF132607	3255 bp	mRNA	PRI	06-MAY-1999
DEFINITION	Homo sapiens histone deacetylase 4 mRNA, complete cds.				
ACCESSION	AF132607				
VERSION	AF132607.1	GI:4754906			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Primates; Catarrhini; Hominoideae; Homo.				
REFERENCE	1 (bases 1 to 3255)				
AUTHORS	Grozinger,C.M., Hassig,C.A. and Schreiber,S.L.				
TITLE	Three proteins define a class of human histone deacetylases related to yeast Hda1p				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	96 (9)			4868-4873 (1999)
MEDLINE	99238449				
REFERENCE	2 (bases 1 to 3255)				
AUTHORS	Grozinger,C.M., Hassig,C.A. and Schreiber,S.L.				
TITLE	Direct Submission				
JOURNAL	Submitted (03-MAR-1999) Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA				
FEATURES	Location/Qualifiers				

[illegible]

Query Match	33.1%;	Score 953.6;	DB 40;	Length 3235;
Best Local Similarity	62.9%;	Pred. No. 5.7e-187;		
Matches 1720;	Conservative	0;	Mismatches 919;	Indels 95;
				Gaps 12;

QY	43	ccctaacccattccctccctccacagacacccccaatgctgtyggagcccaacatgctctctttgga	102
Db	576	CCACGTGATTTTTCAGCGACACCCCTGGCTACGTGTATACGGGAAAGACGACAGATTTCCCTTGA	635
QY	103	ccagatgctccctcccccagagcgcccccttgagagcgtccctctctcaactgacctt	162
Db	636	CCAGAGTTTCTCCACCCCAAGAC-----GGAGTGTGACCTCTTATTAACACCCGGT	686
QY	163	gctctggccctacgacagctcgagacgacttccccctccgcaaaacagctcttgaaccca	222
Db	687	CTTGGGAAATTTAGACAGCCCAAAATGATCTTCCCTTTAGGAAACAGCTTTCGAAACGAA	746
QY	223	ctttaaagtgcttcaaggttaaagaagaagtggtgtgagcggagaagatccctctc	282
Db	747	TCTTAATTAATTCAGGTTCAGGCTTAAGAAGATGGCGGAAACGACGACGCCCCCTT	806
QY	283	gcgtcgcaagatgtagactgttatagcacctttaaagaagagctgtgtgatccag	342
Db	807	ACGGAGGAAGACGGGGCAGTGGTCACTGCTCTATAAAAGCTCCGTTGATGTACAG-	866
QY	343	tgcgggcttggggtcgtgcgtgtgtlaacagcgaccccgctccggccccagctctc	402
Db	866	-----ACTCGCGTGCAGACGCGCCCGGACCTCCGAGCCAGCTCAC	908
QY	403	---aaagatcccaagacacacatcgctgtgagaatgcttactgctcaagtcccaaat	459
Db	909	CACAAACAGCTCCGGGAGCGCTCAGCCCGGAAGAGGATATCGCGCCGCTCCACGAT	968
QY	460	ccccacgagatgtccctcctagacacggagccctccctctgcaacgtcccccacaa	519
Db	969	CCGGCGGAGACAGATTGGCGGACAGACTTGTGGACAGAAAGGCTGGCGGCTCCACT	1028
QY	520	cagctcttacaagtctctctctctctgccaaatctccctaaagtgtgacgccaagtac	579
Db	1029	TCCCTTACACATCGGCATCTTTGCCAAACATCAGCTGGGCGCTGCTGCACCGGCGC-	1088
QY	580	tgtcaccaactcacactcaactcgtcccccgaagctgtgtgacaagcgagggccgagag	639
Db	1088	-----CCTCTGGGGGACAGCGCGGGCCAGCAGACCCGAGAG	1124
QY	640	gcagagccctccagtcctctgcygaggggttggcaagctgcaacggcaagtltcatgagacatc	699
Db	1125	ACGTACCCCTTCCGGCCCTCCAGAGAGAGGCTCCCTTTTCCCGGACCCACCTCATTCC	1184
QY	700	ctctattctctgtgctgtctgtctgtggcggttgagacttgagagggcgagcggaagccccacgggca	759
Db	1185	CTA-----CCTGAGACATCTGGCCCTTGGAGCGGAGCGAGAGGGCA--GCACA	1229
QY	760	tgcctccctgtcagagatgtgtgtgtgtgagagcgccggcgagagagaaacctcat	819
Db	1230	CAGCGCTTCTGTGACGACATGGTCTTACTGAGACGACCCACGACAGACACCTCTGT	1289
QY	820	tgtctgtgcactccacggygaggtcccccacttagtgaaggggtgaagctgtgtgccacaagat	879
Db	1290	CACAGGCTTGGGAGACTGCGCCCTCCACGACAGTCTTGTTGGTGAGAACCGGGGTTC	1349
QY	880	gcgagaggtgtagaaggtcccgcgagcatatcgcccttgagccgacatagctctcaacgct	939
Db	1350	CCCTTCATCCACAAGGTGGGAGACGCGCCACTAGGGGGCGAGCCAGTGGGCCCCCT	1409
QY	940	gcgcgagagtcctccagggcccttgtagcagctgtgtcatgtgacaacaagacacagcaagtctct	999
Db	1410	GCCCCAAAGAGCCCAAGGCTTGTGACGACACTGTGTCACTCAGACGACGATCAAGACTTCT	1469
QY	1000	ggagaaagcaga-----agcagcagcaagctacagctgggcaagatctctcaaca	1047
Db	1470	GGAGAAACACAGAGCAGATTTCCAGCAGACGACAACTCGAGATGAACAGATCATCTCCCA	1529

OY	1048	gaacagggagagctgtcccaaggcagcccccacacccaccccttgaggagagagagagatctac	1107
Db	1530	gccaaagcagccacaccccgccagcccgagacccaccccgaggagacgagagagagacttcgcg	1589
OY	1108	ggaagcagaagaggtctctgtcggggagggagagccctgcacacatgcgcccgagggtccac	1167
Db	1590	tgacacacc---aggccttgcttgagacgagcccttacctgagaccggcttcccgggcacaabga	1646
OY	1168	agagagatgagagacacacaggaagaaccttgagagagagagacgaagaaagagatggygagga	1227
Db	1647	ggccacagcacaagccggcgctgcagagtgaaagcagagcccatggaagacgatgagagaa	1706
OY	1228	ggagagagagatgtcatccaggttaagagacgagagggcgagagatgtgtctcagagagggcc	1287
Db	1707	ggcagagagcccccacacggagagtgagagccggggccacggccacagccagtgagacagagactcct	1766
OY	1288	gcacttgagagga---gcctgggtgtcgtgataacaaaactgttctcagatgtcccaagcgtgtg	1345
Db	1767	ctttcagagacagcaaccccttctcgtctgagagcagcagccagatccacacagcttagaac-----	1822
OY	1346	caaccttgcaggtgtataccaaagcgcccttaagccttgacatgtgtcccccaccaaagccctg	1405
Db	1822	ttaccagggcttcattgagagagggccggcgatccccgtgtcttggggggccacagggcctgtcg	1881
OY	1406	ggccgtataccaaatccctccctgtgtcccttgaggatgaagaaccccccagaacacacc	1465
Db	1882	ttcccccggcgagttcttccaccccgctgcacacttcccgtgtgtgcagagagccccc	1941
OY	1466	gtcagagcaaccttccacacaaagtgtgtctaaagaagttatgtctaagaacagagtc	1522
Db	1942	accaaagccgaggtttcacagacagcgccctgtgtatgacacagctgattgtgaagcacaagtcg	2001
OY	1526	atgtgcggagaaacacacacgtgcacccctgaagatgcgtgcggatcccaagacatcgtgtcc	1585
Db	2002	acctgtcgaggaattacacacagccacaccccgagacagccgggagagattccagagacttctgttc	2061
OY	1586	cgagctgcaagagacagagcctgtctatgaagaatgcagagcgatccagagttgcacaagcag	1645
Db	2062	cgccctggcagagagagggcgcttccggggcaaatgtcgagatctccggcagcgaagccacac	2121
OY	1646	ctagatgagatccagagcaagtgcaactctgaatataccacacccctgtctatgtggagccgtccc	1705
Db	2122	ctggagagagctacagagagcgtgcacttcggaagcccaacccctcctgtatgtgacagaaaccc	2181
OY	1706	ctcaacccgagaaagctcagagacagaaagaattgtctgtgtcccaatcaagcagaagaatgat	1765
Db	2182	cttcaacccggcgaagactggagacagtaagaacattcttagcctcgctgcgc---tccgtgttc	2238
OY	1766	gctgtgtgcctctgttggyggcatccgtgggttgagacagtacacgtyltgaaatgatagtcac	1825
Db	2239	gtmccgggcttcccttgctgggtggttgggggtgagacagtaacacatmtgaaacagagtgac	2298
OY	1826	tcctccagttgtctgtgcacatgacatgagtgagctgcctgcctcagagctgtgcctcaagttgtct	1885
Db	2299	ttcggggggggagagcccccctgtgcctgtgggctgcgtgtagagctgtgttttcmaagctggcc	2358
OY	1886	gcagagagagctcaagaatgattgttcacatataccggcccccagagacacacagccgaagaa	1945
Db	2359	acagagggagctgaagatgtgctgtgtgtccggccccccttgagacacattccgaagag	2418
OY	1946	tcacagagcatgaggaattctgtcttcttaacatctgtatgacataccgcgaanaactctacag	2005
Db	2419	agcaacgcccacatgggcttcttctgtacttcaactccgttgccgtggcagaccgaagtttgcag	2478
OY	2006	cagaagtgaacgttggcacaagttccatctcgtgtgagcttggagaaatcaacatgacatgac	2065
Db	2479	cagaggttgagccgtgagacaagatctctcatctgagatggagagctgacacattgaaacggg	2538
OY	2066	accacagagaggttcttcaaatgtaccctctgtgtctataactctctgtatgtcgtatgac	2125
Db	2539	accacagagaggttcttcaacagacccctgacgctctctgataatgttcccttcacacgctgac	2589
OY	2126	aagcggaactcttccagagctctcgtgggtccctcgaagagtttgttgagaaacagcgtg	2185



[illegible][illegible]





Db	Accession	Source	Organism	Reference	Authors	Title	Journal
QY	2725	gaacagcccaagctcgagccgagccggaacaca	2758				
Db	3931	AACTTCGAGAAAGAAAGAGCCGACGAGCGCACCGCCATGGCCCTCGCTCGCTGGGCGTG	3990				
Db	3991	AAAGCCCGCGAAAGAGACACAGATGAGAGGCCCA	4024				
RESULT	8						
LOCUS	HSM800965						
DEFINITION	Homo sapiens mRNA; CDNA DKFZp586j0917 (from clone DKFZp586j0917); partial cds.						
ACCESSION	U11745						
VERSION	U11745.1	GI:5911906					
KEYWORDS							
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
AUTHORS	1 (bases 1 to 3758)						
TITLE	Koehler, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.						
JOURNAL	Direct Submission						
COMMENT	Submitted (15-SEP-1999) MIPS, Am Klopfersplitz 18a, D-82152 Martinsried, GERMANY						
	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical Research Center at the Charité, Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.						
	This clone is available at the RZPD in Berlin.						
	Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <a href="http://www.mips.biochem.mpg.de/proj/cDNA/">http://www.mips.biochem.mpg.de/proj/cDNA/</a> .						
FEATURES	Location/Qualifiers						
SOURCE	1..3758						
	/organism="Homo sapiens"						
	/db_xref="taxon:9606"						
	/clone="DKFZp586j0917"						
	/clone_id="586 (synonym: hute1). Vector pSport1; host DH10B; sites NotI + SalI/MluI"						
	/dev_strage="adult"						
	/tissue_type="uterus"						
	/map="12q13.1"						
	1..2638						
	/gene="DKFZp586j0917"						
	<1..2638						
	/gene="DKFZp586j0917"						
	/note="strong similarity to histone deacetylase-like protein"						
	/codon_start=2						
	/product="hypothetical protein"						
	/protein_id="CA855935.1"						
	/db_xref="GI:5911907"						
	/translation="SKRSAYASSVYKKALAEVITLKKQALERTVHNSGIPYRTLEPLETEGFRMSLSFLPPVPSLPDPPEHPLKRTVSEPMIKLTKYRKSLERKNPLLRKSSAPSLRRRAEPLIGDSPESSSSSSSSSPNDSSHGPNLIGSBAALLGQRLRLQETSVPALPTVSLPATLTLGLPARADSDGRTPGLGRPLIGSHPTPLFLPLGLPEAGCTPLSRLOPILLDPLSGSHALITVPLGLPPLPHAKOSLWTTERTLSSGLHWPISLRSEPLPSATAPPPGPMQPLRLQTHYVYIKRSAPKSEKPRRLROIASDLEIDGGPGQVDDGLREHLEGHQPARAPALQOHPOVYLMEQQLAARLPREGSTGDSVLLPLAOGGHRPLSRAOSSPAAPASLSAPEASQARVYSSSETPARTLPPTGLIYDVALHKQCSGDNRSRHPREHAGRIQSIWSRJOEGRLSQECLEGRASLEELQSVHSERHVLATGTNPLSLRLKLDNCKLAGILQARMFVMLPCGVGVDDTINNELSHSAARWAAAGVDTAFKVASRELKNGFAVYVRPGHADHSTAMGCFEFSVAIACBLOOOSAAKSKLIVYMDVHNGNGTQOTFYODPSVYILGHHRDGSENFPGSAVDVAGGSEGNVNNAAMAGCLDPPHMGDPEYTLAARIVYVPIIARSPDYIVSAGTDAEGRAPILGCHVSACQETMTQQLMNLNLAGAVVLLAEGBHGLIICDSEACVALLGNRDPILSBEBSWKQPRNLNAIRSLSEAVIRVHSHYCMQRLASCPDSWVRVPGADKEEVAIVTALSLISVGLIAERLPSPEQLVEEPEPNML"						

[illegible]

QY 2435 ctggcccttgagaggagccacttgacccgcatctgtatgctctgaaagctgttc 2494  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 2288 CTGGCCTTGAGGGGTGGCCTAGACCTCAGCCACTCTGTGAGCGCTGTGTG 2347  
QY 2495 tccgctctgctcaagtgtlaaagctgcaagcccttgatgaggaagcttgcaagaaagccc 2554  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 2348 GCTCTCTTCTTCGGGTAACAGGGTGGATCCCTTTCAGAGAAGAGCTGGAACAGAAACCC 2407  
QY 2555 aacatcaagcagatgagcagctagagaagatcatcgatccagagcaaacactgagag 2614  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 2408 AACCTCAATGCATCCGCTCTCTGAGAGCCGCTGATCCGGGTGCACAGATAATCTAGGGGC 2467  
QY 2615 tgtgtgagaagaagtcgcgcgctgctgagccgctccctcgagaggagcccaagcaggtag 2674  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 2468 TGCATGCGAGCGCCTTGCGCTCTGTCCAGACTCTGGGTGCTTAGAGTCCAGGGGCTGAC 2527  
QY 2675 accgaagaagcgaatctgaagcccatgagcctgac 2710  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 2528 AAGAAGAAGTGAAGGCGCAGTAAACCGCACTGGCGTCC 2563

RESULT 9  
AF207749 2871 bp mRNA ROD 07-FEB-2000  
LOCUS Mus musculus histone deacetylase 7 (HDAC7) mRNA, complete cds.  
DEFINITION AF207749  
ACCESSION AF207749  
VERSION AF207749.1 GI:6911183  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 2871)  
AUTHORS Kao, H.Y., Downes, M., Ordentlich, P. and Evans, R.M.  
TITLE Isolation of a novel histone deacetylase reveals that class I and  
class II deacetylases promote SMRT-mediated repression  
Genes Dev. 14 (1), 55-66 (2000)  
JOURNAL 20107033  
MEDLINE 2 (bases 1 to 2871)  
REFERENCE Kao, H.Y., Downes, M., Ordentlich, P. and Evans, R.M.  
AUTHORS Direct Submission  
JOURNAL Submitted (22-NOV-1999) Gene Expression Laboratory, The Salk  
Institute for Biological Studies, 10010 N. Torrey Pines Rd., La  
Jolla, CA 92037, USA

FEATURES  
Source  
1..2871  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="B6:C57BL/6"  
/db\_xref="taxon:10090"  
/dev\_stage="day 14.5 embryo"  
1..2871  
/gene="HDAC7"  
1..2817  
/gene="HDAC7"  
/function="repress transcription"  
/note="class II histone deacetylase"  
/product="histone deacetylase 7"  
/codon\_start=1  
/protein\_id="AAE31419.1"  
/db\_xref="GI:6911184"

gene  
CDS  
1..2817  
/translation="MHSFGACRALQPDPTGSGROPMDLVGQRPVTEPPEPALTL  
QHPDRHLRLAGLHDOORSAEPMRLSMDPMPPELGGQEDLRDLINKDKSKRA  
VASVYKQKLAELYILKQOALERTVSPSPISPYRLIEDLDEGARVLSSELPV  
PSLPTEPPEHPLKRTVSEENLKLRYKPKSLERKNPLRKEBAPSLRRPAETLG  
DSSPSSSTPASGSSPNDESHGNPALGSDADGDRTHSTGLGRGVLAGNPAPLTL  
HGLLEPAGGTLSRLQPIILLDPSVSHALMTYPGIGPLPHHAPDLITTEHSSGS  
LHRLPNTRESEPLPSPATSLAPLADPDRDLKPHYOLIKPAISPPORAKRESEKRL  
LRQIPSAEDLETDGGGVGMANDGLEHRESGRGPEBGRGSIQLQHOQVPWEOHILA  
GRLSQSGSPGDSVLIPLAQVGHRLSRQSSPAAVSNLSSEPTQOTVLSNSETPDS  
LVYDSVMILKQSCGDNKSHPEHAGRIQISIRLOEGRGLSQCCLEGRASLEILOS  
VHSERHVLGTNPLSRKLNDGKLTGLLQAFVPMVPCGAGVDTDTINNELHSSNA  
ARMAAGSVTDLAKFVASRELKNGFAVVRPPGHADHSTAMGFCFENVALACRDLQOH

BASE COUNT 585 a 946 c 801 g 539 t  
ORIGIN  
GKASKLILVMDVHHNGTOOTFOYDPSVLYISLHRHDDGNFPFGSAGVDEVTASGE  
GFNNVNAAGGIDPMDGPEYLAFRIYVWMDIPEAFADLYIVASGDAEAGHPAG  
GYHSACFGYMTQOILMNILAGAVYLLBEGHDIPTAICDSEAKVALLGRKXPLSE  
ESWROKPNLSAIRLEAVRYHRRYWCMQMLASCPDPSWLPRIYVGADAEVAVTALS  
LSVGIILADREPSERLVEEPEPMN."

Query Match 21.7%, Score 625.8; DB 12; Length 2871;  
Best Local Similarity 67.6%; Pred. No. 2,56-119;  
Matches 908; Conservative 0; Mismatches 432; Indels 3; Gaps 2;

QY 1379 ctggcacaatggtgccccaccagaagcccttggtggccgtaccacatctctccctgcgtg -ccccgg 1437  
Db 1413 CCGGCGCTGTCCAGAACCCAGACTCTTCCAGACACACTGTCTTCCATGTTGAGACAGA 1472  
QY 1438 gggcattgaagaacccccagagaccacccgtcaagcactcttcacacaaagtgtgtcta 1497  
Db 1473 GCCACCTGTGAGACCCCAAGTCTCAACAGCTCAGAGACACCTGCTACAGGGCTGTCTA 1532  
QY 1498 cgaacgltcatgctaaagcaccagtgcatgtgcgggaacacacacgltgcacctgagca 1557  
Db 1533 TGACTCGGTGATGCTGAACAACCAATGTTCTGTGAGACAACACAGCAATCCGAGCA 1592  
QY 1558 tgcgtgcccagatccagagcatctgtgtcccggtcgtgagaagaagcctgcttagaagt 1617  
Db 1593 TGCAGGCGGATCCAGACATCTGTGTCCGGCTGAGGAGACGGGGTCTCCGAGCACTG 1652  
QY 1618 cgaagcgatccgaagtgccgaagccagcgtatgatatccagaagaagcactctgaata 1677  
Db 1653 TGAGTGTCTCCGAGGCCCAAAAGCTTCCCTAGAGAGAGCTGCAGTAGTCCACTGTGAACG 1712  
QY 1678 ccacacccctgctctatgagacagtcacctcaaccgcgcgaagaagctagacagcaagaagt 1737  
Db 1713 GCACGTGCTCTTATTCGCGACAGAACCCACTAGCCGCTCAAACTGATGAACGGGAGACT 1772  
QY 1738 gctcgtgtcccatccagcagaagatgtatgtcgtgtcgtcctgtgtgagggatccgggtgga 1797  
Db 1773 TACAGAGACTCTCCGCGACCGGACGTCGTGTGATGTATCCCTGTGGCGGGGTGGTGGA 1832  
QY 1798 cagtgaacagctgtggaatgagatgactcctcctcaagtctgtgagcagtgagtggtctg 1857  
Db 1833 TACTGACACCATCTGGAACGAGACTCATTTCTCCATGACAGCCCTTGCGTGGCGGACG 1892  
QY 1858 cctgctgagcttgagccttcaaggtgtgctgcagagaagctcaagaattggaattgccatcat 1917  
Db 1893 TGTACACGACCTTGCCCTTCAAAAGTAGCTTCCCGAAGAGTGGAAGATGGCTTGTGCTGTG 1952  
QY 1918 ccggaccccaaggacaacagcgccgaggaatccaagacatggatctgtctcttaactc 1977  
Db 1953 GCGAGCCCCGGGACACCATGAGATGAGATCTTAACACCTATGGGCTTCTCTTCAACTC 2012  
QY 1978 tgtagcacaacccgcaaaactcctcacaagcagaagtgtgaagctgaggaagttcctcatcgt 2037  
Db 2013 CGTGGCATTCGCTGCGCCAGACACTTACAGCAACCGGCAAGCAAGATCTCTATTGT 2072  
QY 2038 ggaacttgagacattcaacatlggcaatgacccagcagtcgttctcaaaatgaacccctcgt 2097  
Db 2073 TGACTGGGATGTTCAACATGGGAACGACACAGCAGACTTCTACAGAACCCCAAGTGT 2132  
QY 2098 gctctacaatctctgcatcgctatgacaagggaaactcttccagagctctgaggtccc 2157  
Db 2133 GCTTACATTTTCCCTTCATTCGCAAGAGAGAGCAACTCTTCCAGAGAGTGGGCGCGT 2192  
QY 2158 tgaagaagcttggttgaaggaacagcggtgtgggtatacaattgaacgtgagatgagcagaag 2217  
Db 2193 GGAATAGGTGGGAAGTCCCAAGTGGCGAGAGGCTTCAATGTCAACGATGCGCTTGCGGCG 2252  
QY 2218 tgtgagaccccccatttggaagcgttggaatlaacttcaacagctcaaggaagctgtgtatgct 2277  
Db 2253 CTGTGATCCACCCCAAGGGGAGATCCTGAGTACTGCTGCTTCAAGATAGTGGTATGCC 2312

OY	2278		catggccacgaagctcaccagaatggtgcttaagtcccgcgcgggtttgatgcacttga	2337
Db	2313	CATTGCCCGGAGAGTTTGCTCCAGACCTGGTCTGTGTGTGTTGAATGCTGC	2372	
OY	2338	aggacatctgtctccctcggtgtgctacctctccaacgcccaatgttttgccacttgac	2397	
Db	2373	GGGTCACCACGCCCCCGCTGGGTGGCTTACCATTCTTCTGCAAAATGTTTTGGGTACTGAC	2432	
OY	2398	caggcagctgatgaccttcgcaaggagcgaggtgtgtctgtgccttcgaaaggacata	2457	
Db	2433	GCAGCAGTTGATGAACTTCGCGAGGAGGCGCCCTGGTGTGGCTTAAGAGGCTGGACATCA	2492	
OY	2458	cttagccgcacatctctgtatgtcctctgaagacttgtctctcgcgtcgaagtgtaaagt	2517	
Db	2493	CCTCACGGCCCATCTGTGTATGCTTCGGAAGCCCTGTGTAGCTGCTCTTCTTG6GACAAGA	2552	
OY	2518	gcagcccttgatgtgaggaagctcttcagcacaagaagcccaacatcaacgcagtgccacgt	2577	
Db	2553	GGACCCCTTTTCAGANAAGAGCTGGAACACAGAACCCAAACCTCAGTGGCATCCGGTGGCT	2612	
OY	2578	agagaagaatcagatcagatccagaagcaaacacttgagacttgtgtgcagaagttgcgcgtg	2637	
Db	2613	GGAGCCTGTGGTTCAGGGGTGCACAGGAATACTAGGGGCTGTCAACGACGCGCTTGGCCCTCG	2672	
OY	2638	tctggcgcggtccccctcgsgaggggcccaagcaagtgtagaccagaagaagccgaattgtaac	2697	
Db	2673	TTCACACTCTCTGTGGTACCACAGATCCCGGAGACTCATGTCAGAAATGG - AAACCGTAGCC	2730	
OY	2698	gcacatgacctgtctgtgtgtgtg 2720		
Db	2731	GGCGGGCATCCCTTTCGTGTGG 2753		
RESULT 10				
LOCUS	AC004150	120766 bp	DNA	PRI 17-APR-1999
DEFINITION	Homo sapiens chromosome 17, clone hRRC.97L_F_3, complete sequence.			
ACCESSION	AC004150			
VERSION	AC004150.8	GI:4585948		
KEYWORDS	HTG.			
SOURCE ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE AUTHORS	Bliren,B., Pasmam,K., McKernan,K., Nusbaum,C. and Lander,E. 1 (bases 1 to 120766)			
JOURNAL TITLE	Homo sapiens chromosome 17, clone hRRC.97L_F_3			
REFERENCE AUTHORS	Unpublished 2 (bases 1 to 120766) Bliren,B., Pasmam,K., McKernan,K., Nusbaum,C., Richardson,P., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckeryl,R., Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cerny,J., Cooke,P., Daly,M.J., DePayer,E., Devon,K., Dewar,K., Donelan,L., Durette,B., Flemadi,S., Ferreira,P., Forrest,C., Funke,R., Gage,D., Gardyna,S., Genschlmer,S., Gerlagery,K., Gilmarlin,T., Gray,D., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Linton,L., McKenzie,J., Marguis,N., McEwan,P., McCurk,A., Meldrum,J., Molla,M., Morris,W., Morrow,U., Nachman,A., Naylor,J., O'Connor,T., Paylin,B., Peterson,K., Ranganath,S., Riley,R., Roberts,D., Rollins,G., Rossello,R., Roy,A., Shyam,R., Soohoo,S., Strange-Thomann,N., Stillwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Ye,W.J., Zemseva,I., Zhao,J. and Zody,M.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-FEB-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA			
REFERENCE AUTHORS	3 (bases 1 to 120766) Bliren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckeryl,R., Benn,J., Brown,A., Castle,A., Cerny,J., Collangelo,M., Collins,K., Collymore,A., Cooke,P., DeReitano,K., DePayer,E., Devon,K., Dewar,K., Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,			

TITLE	JOURNAL
<p>Hagos, B., Heatfoot, A., Horton, L., Howland, J.C., Jones, C., Kann, L., Karatas, A., Lehoczek, J., Liu, C., Locke, K., MacDonald, P., Marquis, N., McSwan, P., McGuck, A., McKernan, K., McLaughlin, J., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Ntiloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Strange-Rhmann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Torruella Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, W.</p>	<p>Direct Submission Submitted (17-APR-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Apr 16, 1999 this sequence version replaced gi:4580413. All repeats were identified using RepeatMasker: Smlt, A.F.A. &amp; Green, P. (1996-1997) <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a> location/Qualifiers</p>
source	<p>1. 120766 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="MRPC.971_F_3" /clone_lib="Peter de Jong/ human PAC library" /map="17" /chromosome="17" 1285. .1324 /rpt_family="(CGTG)n" complement(1764. .1935) /rpt_family="MER91A" 8156. .8350 /rpt_family="MIR" 11510. .11540 /rpt_family="(T)n" 11876. .11927 /rpt_family="L2" complement(12664. .12966) /rpt_family="AluSg" complement(13019. .13043) /rpt_family="MIR" 13044. .13334 /rpt_family="AluJo" complement(13335. .13393) /rpt_family="MIR" 13430. .13699 /rpt_family="AluSg" complement(13704. .13872) /rpt_family="MER91A" 13954. .13992 /rpt_family="MER65B" complement(14240. .14426) /rpt_family="MER5A" 15195. .15325 /rpt_family="MIR" 15946. .15960 /rpt_family="Charlie4" complement(15961. .16244) /rpt_family="AluSx" 16235. .16285 /rpt_family="Charlie4" 16286. .16588 /rpt_family="AluSg" 16589. .16661 /rpt_family="Charlie4" complement(16696. .16900) /rpt_family="MIR" 17327. .17631 /rpt_family="AluSg" 17635. .17928 /rpt_family="AluSc" complement(17932. .17987) /rpt_family="MIR" complement(18019. .18188) /rpt_family="MER91A" 18539. .18655</p>

```
repeat_region      /rpt_family="MIR"
complement(18859..19170)
repeat_region      /rpt_family="AluY"
19182..19205
repeat_region      /rpt_family="AT-rich"
19206..19505
repeat_region      /rpt_family="AluSc"
complement(19218..19229)
unsure             /note="Single-stranded terminator coverage."
repeat_region      19506..19524
/rpt_family="AT-rich"
repeat_region      complement(19525..19837)
/rpt_family="AluSg"
19838..19858
repeat_region      /rpt_family="AT-rich"
20517..20822
repeat_region      /rpt_family="AluY"
complement(21146..21433)
repeat_region      /rpt_family="AluSp"
complement(21532..21835)
/rpt_family="AluSx"
repeat_region      23272..23500
/rpt_family="AluSg"
repeat_region      23518..23827
/rpt_family="AluSg"
24479..24589
repeat_region      /rpt_family="C-rich"
24984..25165
repeat_region      /rpt_family="MIR"
25207..25471
repeat_region      /rpt_family="AluUo"
25474..25560
repeat_region      /rpt_family="(TA)n"
complement(25637..25933)
repeat_region      /rpt_family="AluY"
complement(25940..26250)
/rpt_family="AluSx"
26573..26637
repeat_region      /rpt_family="(CAG)n"
26800..26939
repeat_region      /rpt_family="(CAG)n"
complement(29224..29299)
/rpt_family="MER33"
29300..29606
repeat_region      /rpt_family="AluY"
complement(29607..29624)
/rpt_family="MER33"
29625..29908
repeat_region      /rpt_family="AluSg"
complement(29909..29969)
/rpt_family="MER33"
repeat_region      complement(29997..30045)
/rpt_family="AluUo"
complement(30046..30333)
/rpt_family="AluSx"
30222..30254
repeat_region      /note="Single-stranded coverage."
complement(30338..30515)
/rpt_family="AluY"
30338..30358
unsure             /note="Possibly 21 T's, possibly 20 T's."
30366
repeat_region      /note="Probably A, possibly G."
complement(30516..30828)
/rpt_family="AluSg"
repeat_region      complement(30829..30964)
/rpt_family="AluY"
repeat_region      complement(30965..31058)
/rpt_family="AluUo"
complement(31059..31214)
/rpt_family="MER33"
31829..31896
repeat_region      /rpt_family="(CAG)n"
```

```
repeat_region      31991..32280
/rpt_family="AluSg"
repeat_region      32281..32442
/rpt_family="AluSg/x"
repeat_region      32753..32794
/rpt_family="(GGA)n"
repeat_region      33138..33190
/rpt_family="(TG)n"
repeat_region      33285..33315
/rpt_family="(CCCCG)n"
repeat_region      34350..34489
/rpt_family="AluUo"
repeat_region      34490..34784
/rpt_family="AluSx"
repeat_region      34785..34963
/rpt_family="AluUo"
```

Query Match 10.0%; Score 289; DB 40; Length 120766;  
Best Local Similarity 98.3%; Pred. No. 7e-50; Mismatches 5; Indels 0; Gaps 0;  
Matches 292; Conservative 0;

```
QY 1035 agatcctcaccagaacagagagagctgcccagagcagccaccacccctgagagacag 1094
|||||
DB 32593 AGATCCTCACCAGAACAGAGAGAGAGCTGCCAGGCAGCCACCACCTGAGAGACAG 32652
|||||
QY 1095 aggaagagctgacgagacagcagcagaggtcttctgctgggagagagcctgaccatgc 1154
|||||
DB 32653 AGGAGAGAGCTGACGAGAGCGAGAGGAGGCTTCTGGGGAGGAGACCTCGACCATGCCCC 32712
|||||
QY 1155 ggaagagctccacagagagctgagacacacagagaagacctgagagagagagagag 1214
|||||
DB 32713 GGGAGGAGCTCCACAGAGAGTGTGACACACAGAGAGACTGTGAGAGAGAGAGAGAG 32772
|||||
QY 1215 aggaatggagagagagagagagagctgcatccaggttaagagcagagagagagagag 1274
|||||
DB 32773 AGGATGGGAGAGAGAGAGAGAGATGTGATCATCAGAGTTAAGAGAGAGAGAGAGAGTGTG 32832
|||||
QY 1275 ctgaagagagagagagagagagagagagagagagagagagagagagagagagag 1331
|||||
DB 32833 CTGAGGAGAGGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 32889
|||||
```

## RESULT 11

AB018287 4238 bp mRNA PRI 16-JUN-1999  
LOCUS AB018287  
DEFINITION Homo sapiens mRNA for KIAA0744 protein, complete cds.  
ACCESSION AB018287  
VERSION AB018287.1 GI:3882208  
KEYWORDS  
SOURCE Homo sapiens adult male brain cDNA to mRNA, clone.lib:pbluescriptII  
SK plus clone:hk04110.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
AUTHORS Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N.,  
Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.  
TITLE Prediction of the coding sequences of unidentified human genes. XI.  
The complete sequences of 100 new cDNA clones from brain which code  
for large proteins in vitro  
JOURNAL DNA Res. 5 (5), 277-286 (1998)  
MEDLINE 99087487  
REFERENCE 2 (bases 1 to 4238)  
AUTHORS Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.  
TITLE Direct Submission  
JOURNAL Submitted (08-OCT-1998) to the DBJ/EMBL/GenBank databases. Osamu  
Ohara,Kazusa DNA Research Institute, Laboratory of DNA Technology;  
Yana 1532-3, Kiserazu, Chiba 292-0812, Japan  
(E-mail:cdnainfo@kazusa.or.jp, Tel: +81-438-52-3913,  
Fax: +81-438-52-3914)  
FEATURES  
1..4238  
location/Qualifiers  
/organism="Homo sapiens"

/db\_xref="taxon:9606"  
/clone="hK04110"  
/clone\_1lb="PB1uescriptLI SK plus"  
/dev\_stage="adult"  
/sex="male"  
/tissue\_type="brain"  
/gene="K1A0744"  
151..1923  
151..1923  
/gene="K1A0744"  
/codon\_start=1  
/product="K1A0744 protein"  
/protein\_id="BAA34464.1"  
/db\_xref="GI:3882209"  
/translation="MMSISSVDVKSFEVPLSPIDLRDTRMMVPVDPVREK  
GLODFELLIOOOIOIKOLLIAFROKHEMLTROHQAQOEHLKELLAIKOQELLEK  
EKLLEODRODEVERRRRREQLPLPKSKDGRRAVASTVYKLOFELLSKATDY  
PTGNKHSVRKPKLWYTAHHTSLDQSSPLSGTSPSYKTLPGAQADADPLKRT  
ASEPNLKVSRRLKQVAERSSPLLRKDDNVYTSRKMEVTESSVSSSGSGS  
SPNNGPTGAVTENETSVLPPTPHAEQMSQORLIHDSMNLILYTSPLPNTTGL  
PAYVSOLNNSLSKEKQETORLROGVPLPGYGSIPASSHPHYTEGKDPNSSH  
GALLOHLLKEOMROOKILYAGSVPLHPOSPLATKRSIPSGITGKLPRLRNTO  
SAPLPOSTLAQVITQOOHOQFLEKOKOYQOQIHMKILSISIQLKOPSHLEABEE  
LQGDQMOEDRAPSSNSTRSSACVDITLGOVGAVKVEEFPVSDDEDQIOEMESG  
EQAFMOOYIGKDLAPFVTKVII"  
BASE COUNT 1340 a 816 c 850 g 1232 t  
ORIGIN

Query Match 7.78; Score 222.4; DB 9; Length 4238;  
Best Local Similarity 55.7%; Pred. No. 4,4e-36;  
Matches 602; Conservative 0; Mismatches 421; Indels 57; Gaps 7;

QY 55 cctccacagcagcccaaaagctgaggagccaccatgctcttgagcagagttccc 114  
DB 660 CCGCATCCCAAGCTGTGGTACAGCGCTGCCACACACATATTGGATCAAGCTCTCC 719  
QY 115 tccccagagcggccccctggagcagctccctcctaacaactgcttgctggcccta 174  
DB 720 ACCCTTAG-----TGGAAATCTCCATCTCAACAAGTACACATTACAGAGACACA 770  
QY 175 cgcagctgagagagactccccctcgcaaaacagcctctgaaccacactgaagtgcg 234  
DB 771 AGATGCAAGAGATGATTCCCTTCCAAAAATCGCTTGAGGCCCAACTTGAAGTGGCG 830  
QY 235 ttaagagctaaagagagtgctgtagcgagagacagctccctctgctgtagcagga 294  
DB 831 GTCCAGGTTAAACAGAAAGGCGAGAGAGAGAAAGCGCCCTTACTCAGCGCGAAGGA 890  
QY 295 tggagactgataagacacttaagaagagagcgtgtgagatcaagtgccggcgctcg 354  
DB 891 TGGAAATGTTGTCATCTCAATCAAGAAAGCGAATGTTGAGGTGACAGA----- 939  
QY 355 ggcgtgctcgtgtgtaacagcgagcggcgctcgcccccagctctcccaaca---gctc 411  
DB 939 ---ATCTCAAGTCAAGTGCAGGTTCTCCAGGCTCTGTCGCCAGTTCCACCAAAATGGGCGC 995  
QY 412 ccaagacacatcgctgaagatgcttactgctgctagctcgtcccaacaatccccatag-- 470  
DB 996 AACTGGAAGTGTACTGAATAATGAGACTTGGTTTGGCCCTTACCCCTCTATGCGCGAGA 1055  
QY 470 -atgtccctcagcagcagcgcctccctctgagacagctcccccaacagttcagcctcta 528  
DB 1056 AATGCTTTCACAGCAAGCATTTCAATTCATGAGATGATTCATGAACCGTGAATCTTTTA 1115  
QY 529 caagctctcctctctgcaccaacatctccctagagctgcaagcgcaagctactaccca 588  
DB 1116 TACCTCTCTCTTCTTGGCCAAACATTACTTGGGGCTTCCCGCACTG-----CC 1163  
QY 589 ctacacactactgctcccccgaagcgtgtagacacagcagagcgccgagcgagcgct 648  
DB 1164 ATCCAGCTCAATGCTTGATTCATCAAGAAAGCAAGAGTGTAGAG----- 1216

QY 649 ccagctccctgcgagcaggtgtagcagctgaccgcggaattcagacatcctatctcc 708  
DB 1216 -CAACAGCTTAGCGAAGGTGTTCTCTCGCTGGGAGATGAGAGCAGCATCCGGCATC 1274  
QY 709 tgcgtcctctgctgagctgtagcagtgagggcgagcaggaagccccagggcagctcctcc 768  
DB 1275 TTCCAGCCACCTCATGTTACTTTAGAGGGGAACCCACCACAGCAGCAGCGCTCT 1334  
QY 769 gctgcagcatgtgctgtgctgcgagcagccggcagcagagcaccctcatgtc----- 824  
DB 1335 CTTGACGATTTATTATTGAAGAAACAATGCGACAGCAAAAGCTTTGTTAGCTGGTGG 1394  
QY 824 -gtgcacatccagcggagctcccatagtagcaggtggaagctgtggccacacagcagtcgc 882  
DB 1395 AGTTCCCTTACATCCTCAGTCTCCCTTGGCAACAAAGAGAAATTTCACCTGGCATTAG 1454  
QY 883 gacggtgagcagctcccgagctgcgcccctgaagccgacatagctcctcagcgcgc 942  
DB 1455 AGGTACCCACAAATGGCCCTCACAGACCCCTTAACCGAACCAAGTCGACCTTGGC 1514  
QY 943 gcaagatccccaagcagcagctgagctgcatgcaacaacagcagcagcagctccgga 1002  
DB 1515 TCAGAGCAGCTGGGCTAGCTGCTCATTTCAACAGCAACACCAAGCATTTTGGAGAACCA 1574  
QY 1003 gaagcagaagcagcagcagctacagctgaggcaagatcctcacaagaacagggagctgccc 1062  
DB 1575 GAAGCAATTACAGCAGCAGATGCATGACCAAACTGCTTTCGAATCTATTGAACAAC 1634  
QY 1063 cagcagcccccaacccccacccctgagagagagagagagctgacgagcagcagagagt 1122  
DB 1635 GAACCAACAGCAGCAGTCACTTGAGGAAGCAGAGAAAGAGCTTCAGGGGAGCCAGGCGAT 1694

RESULT 12  
AC023855/c  
LOCUS  
DEFINITION  
Homo sapiens chromosome 17 clone RP11-546M21 map 17, LOW-PASS  
SEQUENCE SAMPLING.  
ACCESSION  
AC023855  
VERSION  
AC023855.1 GI:6997304  
KEYWORDS  
HTG; HTGS\_PHASED.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 55893)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 17, clone RP11-546M21  
Unpublished  
2 (bases 1 to 55893)  
Anderson,S., Baldwin,J., Barra,N., Bede,F., Boguslavsky,L.,  
Bouhgalter,B., Brown,A., Burkett,G., Campolano,A., Castle,A.,  
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
Dearlano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,  
Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Riley,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., Landers,T., Laroque,K., Lenocky,J., Levine,R.,  
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,  
McEwan,P., McGuirk,A., McKernan,K., McPeckers,R., Meldrim,J.,  
Meneus,L., Mihova,T., Miranda,C., Menga,Y., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,  
Peterson,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,  
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,  
Travers,M., Trigglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,  
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and  
Zody,M.  
TITLE  
JOURNAL  
Direct Submission  
Submitted (18-FEB-2000) Whitehead Institute/MIT Center for Genome

## COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L6242  
Center clone name: 546\_M\_21  
-----  
\* NOTE: This record contains 72 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
1  
\* 771: contig of 771 bp in length  
\*  
\* 772 1555: contig of 784 bp in length  
\*  
\* 1556 2341: contig of 786 bp in length  
\*  
\* 2342 3121: contig of 780 bp in length  
\*  
\* 3122 3922: contig of 801 bp in length  
\*  
\* 3923 4709: contig of 787 bp in length  
\*  
\* 4710 5472: contig of 763 bp in length  
\*  
\* 5473 6247: contig of 775 bp in length  
\*  
\* 6248 6936: contig of 689 bp in length  
\*  
\* 6937 7707: contig of 771 bp in length  
\*  
\* 7708 8494: contig of 787 bp in length  
\*  
\* 8495 9267: contig of 773 bp in length  
\*  
\* 9268 10035: contig of 768 bp in length  
\*  
\* 10036 10811: contig of 776 bp in length  
\*  
\* 10812 11565: contig of 754 bp in length  
\*  
\* 11566 12332: contig of 767 bp in length  
\*  
\* 12333 13050: contig of 718 bp in length  
\*  
\* 13051 13838: contig of 788 bp in length  
\*  
\* 13839 14633: contig of 795 bp in length  
\*  
\* 14634 15403: contig of 770 bp in length  
\*  
\* 15404 16188: contig of 785 bp in length  
\*  
\* 16189 16964: contig of 776 bp in length  
\*  
\* 16965 17781: contig of 817 bp in length  
\*  
\* 17782 18561: contig of 780 bp in length  
\*  
\* 18562 19335: contig of 774 bp in length

\* 19336 20112: contig of 777 bp in length  
\* gap of unknown length  
\*  
\* 20113 20888: contig of 776 bp in length  
\* gap of unknown length  
\*  
\* 20889 21651: contig of 763 bp in length  
\* gap of unknown length  
\*  
\* 21652 22419: contig of 768 bp in length  
\* gap of unknown length  
\*  
\* 22420 23208: contig of 789 bp in length  
\* gap of unknown length  
\*  
\* 23209 23992: contig of 784 bp in length  
\* gap of unknown length  
\*  
\* 23993 24768: contig of 776 bp in length  
\* gap of unknown length  
\*  
\* 24769 25539: contig of 771 bp in length  
\* gap of unknown length  
\*  
\* 25540 26302: contig of 763 bp in length  
\* gap of unknown length  
\*  
\* 26303 27084: contig of 782 bp in length  
\* gap of unknown length  
\*  
\* 27085 27876: contig of 792 bp in length  
\* gap of unknown length  
\*  
\* 27877 28636: contig of 760 bp in length  
\* gap of unknown length  
\*  
\* 28637 29419: contig of 783 bp in length  
\* gap of unknown length  
\*  
\* 29420 30186: contig of 767 bp in length  
\* gap of unknown length  
\*  
\* 30187 30964: contig of 778 bp in length  
\* gap of unknown length  
\*  
\* 30965 31748: contig of 784 bp in length  
\* gap of unknown length  
\*  
\* 31749 32537: contig of 789 bp in length  
\* gap of unknown length  
\*  
\* 32538 33313: contig of 776 bp in length  
\* gap of unknown length  
\*  
\* 33314 34089: contig of 776 bp in length  
\* gap of unknown length  
\*  
\* 34090 34882: contig of 793 bp in length  
\* gap of unknown length  
\*  
\* 34883 35680: contig of 798 bp in length  
\* gap of unknown length  
\*  
\* 35681 36450: contig of 770 bp in length  
\* gap of unknown length  
\*  
\* 36451 37224: contig of 774 bp in length  
\* gap of unknown length  
\*  
\* 37225 37997: contig of 773 bp in length  
\* gap of unknown length  
\*  
\* 37998 38778: contig of 781 bp in length  
\* gap of unknown length  
\*  
\* 38779 39553: contig of 775 bp in length  
\* gap of unknown length  
\*  
\* 39554 40319: contig of 766 bp in length  
\* gap of unknown length  
\*  
\* 40320 41098: contig of 779 bp in length  
\* gap of unknown length  
\*  
\* 41099 41890: contig of 792 bp in length  
\* gap of unknown length  
\*  
\* 41891 42681: contig of 791 bp in length  
\* gap of unknown length  
\*  
\* 42682 43440: contig of 759 bp in length  
\* gap of unknown length  
\*  
\* 43441 44208: contig of 768 bp in length  
\* gap of unknown length  
\*  
\* 44209 44971: contig of 763 bp in length  
\* gap of unknown length  
\*  
\* 44972 45757: contig of 786 bp in length  
\* gap of unknown length  
\*  
\* 45758 46538: contig of 781 bp in length  
\* gap of unknown length  
\*  
\* 46539 47304: contig of 766 bp in length  
\* gap of unknown length





OY	2396	accgaagcgcgtgacgacctggcaaggccggcgctggtgtgtctgcccttggaaaggaagccat	2455	
Dd	6007	ACCCCGGCAATCTTTCAGTTGGCCAAAGGTGCCTGCCCTTCGAAGGGCGGTACC	5948	
OY	2456	gacttgaccgcgatcttgtgagtccctctgaactctgtgtcttcgtcgtctgaagtgtaag	2515	
Dd	5947	GATTGCGCCCCCATTTTGATTTCCGCCAAGAAGTGTGTGCGGCGCCTGCTCGGAGATCCC	5888	
OY	2516	ctgcagcccccttgatgatgtagcagctcttcgacgaagaagcccacatcacagcagttggccaacg	2575	
Dd	5887	GOTGTCCTCCATTGGCCAAGGCCGAGCTGGAGGCCGCCCTTCGCAGAATGCCATCATTAACG	5828	
OY	2576	ctgagagaatcatcgatcgatccagagcaa	2604	
Dd	5827	CTCCAGAAGACGATGACCATACAGGTAAA	5799	
RESULT 14				
LOCUS	ACO10993/c			
DEFINITION	ACO10993 182015 bp DNA HTG 15-FEB-2000			
ACCESSION	ACO10993			
VERSION	ACO10993.9 GI:6978352			
KEYWORDS	HTG; HTGS; PHASEL.			
SOURCE	fruit fly.			
ORGANISM	Drosophila melanogaster			
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
AUTHORS	1 (bases 1 to 182015) Celisner,S.E., Agbayani,A., Arcaluna,T.T., Baxter,E., Blazej,R.G., Burenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Richards,S., Sethi,H., Svirskas,R.R., Man,K.H., Webster,D., Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.			
TITLE	JOURNAL. SEQUENCING OF DROSOPHILA MELANOASTER UNPUBLISHED			
COMMENT	2 (bases 1 to 182015) Celisner,S.E., Agbayani,A., Arcaluna,T.T., Baxter,E., Blazej,R.G., Burenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snlt,E., Svirskas,R.R., Man,K.H., Weinburg,T., Zhang,R., Zleran,L.L. and Rubin,G.M.  Direct Submission Submitted (29-SEP-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA On Feb 15, 2000 this sequence version replaced g1:6838492. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site ( <a href="http://www.fruitfly.org/sequence/">http://www.fruitfly.org/sequence/</a> ) or send email to bug@fruitfly.berkeley.edu. All contigs in this submission meet the following cutoffs: length >= 200 bases. NOTE: This is a 'working draft' sequence. It currently consists of 135 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.			
*	1	456:	contig of 456 bp in length	
*	457	536:	gap of unknown length	
*	537	750:	contig of 214 bp in length	
*	751	830:	gap of unknown length	
*				
*	831	1216:	contig of 386 bp in length	
*	1297	1296:	gap of unknown length	
*	1297	1959:	contig of 663 bp in length	
*	1960	2039:	gap of unknown length	
*	2040	2473:	contig of 434 bp in length	
*	2474	2553:	gap of unknown length	
*	2554	3148:	contig of 515 bp in length	
*	3069	3680:	contig of 532 bp in length	
*	3149	3760:	gap of unknown length	
*	3681	4064:	contig of 304 bp in length	
*	4065	4144:	gap of unknown length	
*	4145	4586:	contig of 442 bp in length	
*	4587	4666:	gap of unknown length	
*	4667	5114:	contig of 448 bp in length	
*	5115	5194:	gap of unknown length	
*	5195	5689:	contig of 495 bp in length	
*	5690	5769:	gap of unknown length	
*	5770	6330:	contig of 561 bp in length	
*	6331	6410:	gap of unknown length	
*	6411	7118:	contig of 708 bp in length	
*	7119	7198:	gap of unknown length	
*	7199	7811:	contig of 613 bp in length	
*	7812	7891:	gap of unknown length	
*	7892	8703:	contig of 812 bp in length	
*	8704	8783:	gap of unknown length	
*	8784	9264:	contig of 481 bp in length	
*	9265	9345:	gap of unknown length	
*	9345	9763:	contig of 419 bp in length	
*	9764	9843:	gap of unknown length	
*	10763	10763:	contig of 920 bp in length	
*	10764	10843:	gap of unknown length	
*	10844	11263:	contig of 420 bp in length	
*	11264	11343:	gap of unknown length	
*	11344	11810:	contig of 467 bp in length	
*	11811	11890:	gap of unknown length	
*	11891	12486:	contig of 596 bp in length	
*	12487	12566:	gap of unknown length	
*	12567	13756:	contig of 1190 bp in length	
*	13757	13836:	gap of unknown length	
*	13837	14335:	contig of 499 bp in length	
*	14336	14415:	gap of unknown length	
*	14416	15145:	contig of 730 bp in length	
*	15146	15225:	gap of unknown length	
*	15226	15946:	contig of 721 bp in length	
*	15947	16026:	gap of unknown length	
*	16027	17048:	contig of 1022 bp in length	
*	17049	17128:	gap of unknown length	
*	17129	18037:	contig of 909 bp in length	
*	18038	18117:	gap of unknown length	
*	18118	18671:	contig of 554 bp in length	
*	18672	18751:	gap of unknown length	
*	18752	19606:	contig of 855 bp in length	
*	19607	19686:	gap of unknown length	
*	19687	20382:	contig of 616 bp in length	
*	20383	20382:	gap of unknown length	
*	20383	21257:	contig of 795 bp in length	
*	21178	21257:	gap of unknown length	
*	21258	21650:	contig of 393 bp in length	
*	21651	21730:	gap of unknown length	
*	21731	22910:	contig of 1100 bp in length	
*	22911	23699:	contig of 759 bp in length	
*	23670	23749:	gap of unknown length	
*	23750	24686:	contig of 937 bp in length	
*	24687	24766:	gap of unknown length	
*	24767	25580:	contig of 1214 bp in length	
*	25581	26060:	gap of unknown length	
*	26061	27037:	contig of 977 bp in length	
*	27038	27117:	gap of unknown length	
*	27118	28678:	contig of 1561 bp in length	
*	28679	28758:	gap of unknown length	
*	28759	30049:	contig of 1291 bp in length	

[illegible]

[illegible][illegible]

Search completed: May 20, 2000, 00:40:36  
Job time: 13040 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 00:40:36 ; Search time 6491.04 Seconds

(without alignments)  
-194.527 Million cell updates/sec

Title: US-09-502-945-3  
Perfect score: 1298  
Sequence: 1 ggctgctgaatgactgcga.....gaagaagaagaagaaaa 1298

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 segs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pr5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1288	99.2	1298	40 AF039694	AF039694 Homo sapi
2	729	56.2	5585	9 D87455	D87455 Human mRNA
3	729	56.2	208577	33 AL135082	AL135082 Homo sapi
4	308.6	23.8	472	13 G22194	G22194 human STS W
5	215.4	16.6	7218	5 I66494	I66494 Sequence 14
6	190.8	14.7	131025	52 AC022324	AC022324 Homo sapi
7	189.6	14.6	53156	44 AC021195	AC021195 Homo sapi
8	182.8	14.1	129404	54 AC013349	AC013349 Homo sapi
9	182.4	14.1	198664	44 AC010649	AC010649 Homo sapi
10	181.4	14.0	155284	41 AC009524	AC009524 Homo sapi
11	180	13.9	76897	42 AC016179	AC016179 Homo sapi
12	179.8	13.9	40482	11 HSAC000361	HSAC000361 Human cos
13	179.8	13.9	40977	11 HSAC000362	HSAC000362 Human cos
14	178.4	13.8	43346	54 AC011556	AC011556 Homo sapi
15	178.2	13.7	62015	55 AC023326	AC023326 Homo sapi
16	176.8	13.6	180385	40 AC007461	AC007461 Homo sapi
17	176.2	13.6	42141	34 CEY54G9A	AL032648 Caenorhab
18	176.2	13.6	300197	32 CEY54G9A	298869 Caenorhabd
19	175.8	13.5	140350	43 AC015830	AC015830 Homo sapi
20	175	13.5	227070	56 AC023888	AC023888 Homo sapi
21	174.4	13.4	179285	32 CNS01DS6	AL121656 Homo sapi
22	174	13.4	879	13 CNS01JRG	AL147405 Anopheles
23	174	13.4	73455	45 AC021589	AC021589 Homo sapi
24	173.2	13.3	73020	54 AC022851	AC022851 Homo sapi
25	172.2	13.3	68746	44 AC021008	AC021008 Mus muscu
26	172	13.3	169424	42 AC011961	AC011961 Homo sapi
27	171.2	13.2	120007	11 AF064864	AF064864 Homo sapi
28	171.2	13.2	120609	40 AF165176	AF165176 Homo sapi
29	170.8	13.2	175374	45 AC019265	AC019265 Homo sapi
30	169.6	13.1	59815	44 AC019096	AC019096 Homo sapi
31	169.2	13.0	129149	11 HS10D11	298044 Human DNA S
32	168.2	13.0	175132	40 AC008134	AC008134 Homo sapi
33	168.2	13.0	123132	52 AC013455	AC013455 Homo sapi
34	168	12.9	162575	33 AC004086	AC004086 Homo sapi
35	167.8	12.9	175630	56 AC023059	AC023059 Mus muscu
36	166.6	12.8	93615	40 AC004950	AC004950 Homo sapi
37	166.6	12.8	142257	43 AC017491	AC017491 Drosophi
38	166	12.8	190007	55 AC013314	AC013314 Homo sapi
39	166	12.8	190538	33 AL135926	AL135926 Homo sapi
40	165.2	12.7	34731	11 HS075931	U75931 Human Xp22
41	165.2	12.7	99235	43 AC017010	AC017010 Homo sapi
42	164.2	12.7	127778	44 AC020735	AC020735 Homo sapi
43	164	12.6	220657	52 AC007500	AC007500 Homo sapi
44	163.8	12.6	241350	58 AC021663	AC021663 Homo sapi
45	163.2	12.6	247535	42 AC012645	AC012645 Homo sapi

ALIGNMENTS

RESULT 1  
AF039694 1298 bp mRNA PRI 05-MAY-1999  
LOCUS Homo sapiens antigen NY-CO-16 mRNA, complete cds.  
DEFINITION AF039694  
VERSION AF039694.1 GI:3170187  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (bases 1 to 1298)  
Scanlan,M.J., Chen,Y.T., Williamson,B., Gure,A.O., Stockert,E.,  
Gordan,J.D., Tureci,O., Sahin,U., Pfeundscher,M. and Old,L.J.  
TITLE Characterization of human colon cancer antigens recognized by  
autologous antibodies  
JOURNAL Int. J. Cancer 76 (5), 652-658 (1998)  
MEDLINE 98272252  
REFERENCE 2 (bases 1 to 1298)  
Scanlan,M.J.  
AUTHORS Direct Submission  
TITLE Submitted (23-DEC-1997) Ludwig Institute for Cancer Research, New  
York Branch at Memorial Sloan-Kettering Cancer Center, 1275 York  
Avenue, New York, NY 10021, USA  
FEATURES  
source location/Qualifiers  
1..1298  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="colorectal carcinoma"  
11..406  
/codon\_start=1  
/product="antigen NY-CO-16"  
/protein\_id="AAC18043.1"  
/db\_xref="GI:3170188"  
/translation="MTANRLAESLIALSHOEFLADLPKDYLLSEDEGNDGERKHQ  
KLEFAISLDGKNRKLAESEASLKYSEFNVSSESGEKLVLADLLEPKVTSLSLAT  
VKKRLSIVKSKVLEPLNKEIERIHRE"  
BASE COUNT 548 a 190 c 381 g 174 t 5 others  
ORIGIN

Query Match 99.2%; Score 1288; DB 40; Length 1298;  
Best Local Similarity 99.6%; Pred. No. 3.5e-210;  
Matches 1293; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggctgcgtgaatgactgcgaacccgcttcgacagagccttcgtcttgagccancagga 60  
Db 1 GGCtGCTGAAATGACTGCGAACCGCTTGACAGAGCCTTCTGGCTTGACCCACACAGA 60  
QY 61 agaactagcggatttcgcaaaaagactacctcttgagtgagagtgagaatgagggagaca 120  
Db 61 AGAaCTAGCGGATTTCGCAAAAAGACTACCTCTTGAGTGAAGTGAAGGAGGAGACA 120  
QY 121 tgatcgagagagaaagcatcnaaagctcttgaaagcaatcagcttccttga tggaaagaa 180  
Db 121 TGATGAGAGAGAGAAAGCATCAAAAGCTCTGGAAGCAATCAGTCCCTTGATGAAAGAA 180  
QY 181 tagcgagaaatgctgctgagagtgctgagctagctcagagtgctcagagtgctcaatgtcag 240  
Db 181 TAGCGAGGAAATGCTGAGAGCTGAGGCTAGTCTGAAGGTGTCAGAGTTCAATGTCTAG 240  
QY 241 ttctgaagagatcagaagaaagctgctccttcgacagatcgtcttgaagcctgttaaaacttc 300  
Db 241 TTCTGAAGAGATCAGAGAAAGCTGCTGCTTGCAGATCTGCTTGAGCCTGTTAAAACTTC 300  
QY 301 atctctcttgccactgtgaaaaagcaactagtagagtcanaatcaaagaanaacagtgga 360  
Db 301 ATCTTCTTTGGCCACTGTGAAAAAGCAACTGAGTAGAGTCAAAATCAAGAAGACAGTGA 360  
QY 361 gttacctctgacaagaagagatggaaggtccacagagagaatgacatcaataaagc 420  
Db 361 GTTACCTCTGAAACAAGAGATTGAAGCGATCCACAGAGAAATAGCATTCATTAACG 420

QY 421 cacaagtcctcccaaatgaggaccctgtcgtcctgaaagaaccggcagagcagctg 480  
Db 421 CACAAGTCTCTCCAAATGGAGACCTGTGTCCTCGAAGACCGGACAGCAAGCTGG 480  
QY 481 ttctcccttgagaaagagagcagccatgtctcccatgtgaacatgtgtcagtgct 540  
Db 481 TTTTCCCTCGAGAAAGAGGAGCCAGCATGTGCTCCCATTAACATGTGCTCAGTGCT 540  
QY 541 ggaagcgcaagatccctccttgagcaggaatcttccaactcttcataagaacaagcagc 600  
Db 541 GGAAGCGCAAGATCCCTCTGAGCAGGAAATTTTCAACCTCTCCATAAAGAACAGACG 600  
QY 601 cagtgcacagacccttaactgacccctgtgaaaagcctcctcctcgagccatgagctag 660  
Db 601 CAGTGACAGACCTTTACTGACCCCTGTGAAAAAGCCTCTCTCGAGCCATGAGCCTAG 660  
QY 661 aagagcaagaatgcagacagcagagcttcagaggtctcgggtctcgcagtcctactatg 720  
Db 661 AAGAGCCAAAGATGCGACGACAGACAGCTTCAGAGGCTCGGGCTCTGCGAGCTACTATG 720  
QY 721 angccaaggtctcgaagagagagaataatcnaaagttaaagtatccacaagtcgtgaaga 780  
Db 721 ANGCCAAAGCTCGAAGAGAGAAATAATCNAAGTTAAAGTATCCACAAAGCTGTGAAGA 780  
QY 781 aaggaagcgccaagaaagccctaaagagtttgagcaagctgcggaaggttaatccagctg 840  
Db 781 AAGGAAGGCCAAAGAAAGCCCTAAAGAGTTTGAGCAAGCTCGGAAGGTTAATCCAGCTG 840  
QY 841 ccgcactagaaacgaaagaaagagaaagagagagagagagaaagaaagaaagaaag 900  
Db 841 CCGCACTAGAAAGAAAGAAAGAGAAAGAGAAAGAGAGAGAGAGAGAAAGAAAGAAAG 900  
QY 901 agaaagaaagaaag 960  
Db 901 AGAAGAAAGAAAGAAAG 960  
QY 961 gagaagaaag 1020  
Db 961 GAGAAGAAAG 1020  
QY 1021 agaggaagaaatagaaagagaaagaaagaaagaaagaaagaaagaaagaaagaa 1080  
Db 1021 AGAGGAAGAAATTAAG 1080  
QY 1081 aggaagaaag 1140  
Db 1081 AGGAAG 1140  
QY 1141 aacnagaag 1200  
Db 1141 AACNAGAAAG 1200  
QY 1201 ggaagaaag 1260  
Db 1201 GGAAGAAAG 1260  
QY 1261 agaaagtataagaaag 1298  
Db 1261 AGAAAGTATAAG 1298

RESULT 2  
DB7455 5885 bp mRNA PRI 10-JUL-1997  
LOCUS Human mRNA for KIAA0266 gene, complete cds.  
DEFINITION DB7455  
ACCESSION DB7455.1 GI:1665798  
VERSION DB7455.1  
KEYWORDS KIAA0266  
SOURCE Homo sapiens male bone marrow myeloblast cell line:KG-1 cDNA to  
mRNA, clone\_id:pb1uescript II SK clone:BA2755.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

REFERENCE	ADTHORS	JOURNAL	TITLE	REFERENCE	ADTHORS	JOURNAL	TITLE	FEATURES	SOURCE
1	(bases 1 to 5585)	Primates; Catarrhini; Hominiidae; Homo.							
2	(sites)	Nagase,T., Seki,N., Ishikawa,K. and Nomura,N.							
3	(sites)	Nagase,T., Seki,N., Ishikawa,K., Ohira,M., Kawarabayashi,Y., Ohara,O., Tanaka,A., Kotani,H., Miyajima,N. and Nomura,N.							
4		Ohara,O., Tanaka,A., Kotani,H., Miyajima,N. and Nomura,N.							
5		Nomura, Kazusa DNA Research Institute, Gene Structure 1; 1532-3							
6		Yana, Kizarazu, Chiba 292, Japan (E-mail:cdna1nfo@kazusa.or.jp, URL:http://www.kazusa.or.jp, Tel.0438-52-3930, Fax:0438-52-3931)							
7		analysis of cDNA clones from human cell line KG-1 and brain							
8		analysis of cDNA clones from cell line KG-1 and brain							
9		analysis of cDNA clones from cell line KG-1 and brain							
10		analysis of cDNA clones from cell line KG-1 and brain							
11		analysis of cDNA clones from cell line KG-1 and brain							
12		analysis of cDNA clones from cell line KG-1 and brain							
13		analysis of cDNA clones from cell line KG-1 and brain							
14		analysis of cDNA clones from cell line KG-1 and brain							
15		analysis of cDNA clones from cell line KG-1 and brain							
16		analysis of cDNA clones from cell line KG-1 and brain							
17		analysis of cDNA clones from cell line KG-1 and brain							
18		analysis of cDNA clones from cell line KG-1 and brain							
19		analysis of cDNA clones from cell line KG-1 and brain							
20		analysis of cDNA clones from cell line KG-1 and brain							
21		analysis of cDNA clones from cell line KG-1 and brain							
22		analysis of cDNA clones from cell line KG-1 and brain							
23		analysis of cDNA clones from cell line KG-1 and brain							
24		analysis of cDNA clones from cell line KG-1 and brain							
25		analysis of cDNA clones from cell line KG-1 and brain							
26		analysis of cDNA clones from cell line KG-1 and brain							
27		analysis of cDNA clones from cell line KG-1 and brain							
28		analysis of cDNA clones from cell line KG-1 and brain							
29		analysis of cDNA clones from cell line KG-1 and brain							
30		analysis of cDNA clones from cell line KG-1 and brain							
31		analysis of cDNA clones from cell line KG-1 and brain							
32		analysis of cDNA clones from cell line KG-1 and brain							
33		analysis of cDNA clones from cell line KG-1 and brain							
34		analysis of cDNA clones from cell line KG-1 and brain							
35		analysis of cDNA clones from cell line KG-1 and brain							
36		analysis of cDNA clones from cell line KG-1 and brain							
37		analysis of cDNA clones from cell line KG-1 and brain							
38		analysis of cDNA clones from cell line KG-1 and brain							
39		analysis of cDNA clones from cell line KG-1 and brain							
40		analysis of cDNA clones from cell line KG-1 and brain							
41		analysis of cDNA clones from cell line KG-1 and brain							
42		analysis of cDNA clones from cell line KG-1 and brain							
43		analysis of cDNA clones from cell line KG-1 and brain							
44		analysis of cDNA clones from cell line KG-1 and brain							
45		analysis of cDNA clones from cell line KG-1 and brain							
46		analysis of cDNA clones from cell line KG-1 and brain							
47		analysis of cDNA clones from cell line KG-1 and brain							
48		analysis of cDNA clones from cell line KG-1 and brain							
49		analysis of cDNA clones from cell line KG-1 and brain							
50		analysis of cDNA clones from cell line KG-1 and brain							
51		analysis of cDNA clones from cell line KG-1 and brain							
52		analysis of cDNA clones from cell line KG-1 and brain							
53		analysis of cDNA clones from cell line KG-1 and brain							

Db	841	TGATGAGAGAGAAAGCATCAAAAAGCTTCTGGAGCAATCAATTCCCTGTATGGAAGAA	900
Qy	181	taagcggaaatcgtgctganaagtcctgaagcctagctgaaagtgctcaagtgctcaatgctcag	240
Db	901	TAGCGGAAATTTGGCTGAGAGGCTCTGAGGCTGATGTGAAGTGTACAGATTCAAGTGTGAG	960
Qy	241	ttcgaagagctcaaggaagaagctgctgcttcagatcgcctgagcctgttaaaactc	300
Db	961	TTTCGAAGGATCAGGAAAAAGCTGGGCTTGCAGATTCGCTTGAGCCGTTAAAACTTC	1020
Qy	301	atctctcttgccactctgnaaaaagaactgagtagagtcanaatcaagaanaacagctgta	360
Db	1021	ATCTCTTGTGGCCACTGTAAAAAGCACTGAAATAGAGTCAATCAAAAGAAAGTGTGTGA	1080
Qy	361	gttaacctctgaacaagaagaagatctgaacggtatccacagagaa-tagcatccaataaa-	419
Db	1081	GTTTACCTCTTTAAAGAAAAAATTTGAAACAGATTCACAGAGAAAGTATTCAGTTCAGTAAAC	1140
Qy	419	cggcaagctctctccaatgggaacctgtgcttcctggaagacggcgaggaagcagact	478
Db	1141	CTCACAGTCTCTCCCAAATGGGAACCTTATCATCTCGAAGAACCGACGAGCAGACACT	1200
Qy	479	gatttccctctgagaagaagagagcaagccattgctccatttgatactgtcctaagcgg	538
Db	1201	GGTTTTTCCCTTGGGGAAGAGACAGCAGCCATGCTCCCATTTGAACATCGCCTCAGTGG	1260
Qy	539	ctggaagcagaagactccctctgagcaggaagaattttcaacctctccataaagaacaaga	598
Db	1261	CTGGAAGGCAAGAACTCCCTCTGGAGCAGGAATTTTAACTCTCTCAATGAAGAACAGCA	1320
Qy	599	gcacagtacagacccttctactgacccctgtgtaaaaaagcctctctccgagccatgagcct	658
Db	1321	GCCAGTACAGATCTTTACTGTGACTCCCATGTGMAAAGGCGCTCTCTCCAAAGCCATGAGCT	1380
Qy	659	agaagagcgaagaatgagcagcagagacttcaagaagctctgaggtctctgcagctcacta	718
Db	1381	GGAAGAGGCAAGATGACCGACAGAGCTTCAGAGGCTCGGGCTCTGCAGTCTCTACTA	1440
Qy	719	tgangcgaaggtctcgaaagagagaagaanaatcnaaagtttaaaagltacaaagtcgtgaa	778
Db	1441	TGAGGCCAAGGCTCGAATAAGAGAAATTCAAAAGTATAAAAAGTATCAAAAGTGTGAA	1500
Qy	779	gaaaggaagggccaagaagaagccttaaaagatgttgacagctgcggaaggttaatccagc	838
Db	1501	GAAGGAAAGGCGCAAGAAACCTTTAAAGAGTTTGAGCGCTACAGAAAGTTAAATCCAC	1560
Qy	839	tgcgcagctgagaaagaagaagaagaaggaaggaaggaaggaaggaagaagaagaaca	898
Db	1561	TGTGGCACTGGAGAAATGTGAAAATTTGAAAATGCCCAAAATGATGGAAGAAAGATGAGCT	1620
Qy	899	ggagaagaagaa 910	
Db	1621	TAAACCAACAAA 1632	
RESULT 3			
LOCUS	AL139082		
DEFINITION	AL139082 208577 bp DNA	HTG	17-FEB-2000
ACCESSION	AL139082		
VERSION	AL139082.3	GI:6996171	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Enkaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homini; Hominidae; Homo.		
TITLE	1 (bases 1 to 208577)		
JOURNAL	Submitted (17-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,		
	CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk		
	request: clonequest@sanger.ac.uk		

## COMMENT

On Feb 17, 2000 this sequence version replaced gi:6996045.  
IMPORTANT: This sequence is unfinished and does not necessarily  
represent the correct sequence. Work on the sequence is in  
progress and the release of this data is based on the understanding  
that the sequence may change as work continues. The sequence may  
be contaminated with foreign sequence from E.coli, yeast, vector,  
phage etc. Order of segments is not known; 800 n's separate  
segments. Contig\_ID: 00061 Length: 1830bp

Contig\_ID: 00142 Length: 4909bp  
Contig\_ID: 00300 Length: 22197bp  
Contig\_ID: 00329 Length: 5487bp  
Contig\_ID: 00405 Length: 3512bp  
Contig\_ID: 00474 Length: 15770bp  
Contig\_ID: 00570 Length: 1060bp  
Contig\_ID: 00764 Length: 1246bp  
Contig\_ID: 00787 Length: 21288bp  
Contig\_ID: 00809 Length: 38984bp  
Contig\_ID: 00821 Length: 7384bp  
Contig\_ID: 00873 Length: 14235bp  
Contig\_ID: 01015 Length: 4848bp  
Contig\_ID: 01108 Length: 11783bp  
Contig\_ID: 01128 Length: 11529bp  
Contig\_ID: 01245 Length: 2609bp  
Contig\_ID: 01300 Length: 3272bp  
Contig\_ID: 01543 Length: 12772bp  
Contig\_ID: 01635 Length: 5439bp  
Contig\_ID: 01693 Length: 1408bp  
Contig\_ID: 01711 Length: 1015bp

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 21 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1830: contig of 1830 bp in length  
\* 1831 2630: gap of 800 bp  
\* 2631 7539: contig of 4909 bp in length  
\* 7540 8339: gap of 800 bp  
\* 8340 30536: contig of 22197 bp in length  
\* 30537 31336: gap of 800 bp  
\* 31337 36823: contig of 5487 bp in length  
\* 36824 37623: gap of 800 bp  
\* 37624 41135: contig of 3512 bp in length  
\* 41136 41935: gap of 800 bp  
\* 41936 57705: contig of 15770 bp in length  
\* 57706 58505: gap of 800 bp  
\* 58506 59565: contig of 1060 bp in length  
\* 59566 60365: gap of 800 bp  
\* 60366 61611: contig of 1246 bp in length  
\* 61612 62411: gap of 800 bp  
\* 62412 83699: contig of 21288 bp in length  
\* 83700 84499: gap of 800 bp  
\* 84500 123483: contig of 38984 bp in length  
\* 123484 124283: gap of 800 bp  
\* 124284 131667: contig of 7384 bp in length  
\* 131668 132467: gap of 800 bp  
\* 132468 146702: contig of 14235 bp in length  
\* 146703 147502: gap of 800 bp  
\* 147503 152350: contig of 4848 bp in length  
\* 152351 153150: gap of 800 bp  
\* 153151 164933: contig of 11783 bp in length  
\* 164934 165733: gap of 800 bp  
\* 165734 177262: contig of 11529 bp in length  
\* 177263 178062: gap of 800 bp  
\* 178063 180671: contig of 2609 bp in length  
\* 180672 181471: gap of 800 bp  
\* 181472 184743: contig of 3272 bp in length  
\* 184744 185843: gap of 800 bp  
\* 185844 196315: contig of 12772 bp in length  
\* 196316 199115: gap of 800 bp  
\* 199116 204554: contig of 5439 bp in length

\* 204555 205354: gap of 800 bp  
\* 205355 206762: contig of 1408 bp in length  
\* 206763 207562: gap of 800 bp  
\* 207563 208577: contig of 1015 bp in length.  
Location/Qualifiers  
1. .208577  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/map="q14.3-21.1"  
/clone="RP11-24865"  
/clone\_lib="RPC1-11.1"  
BASE COUNT 58247 a 39896 c 39589 g 54792 t 16053 others  
ORIGIN

Query Match 56.2%; Score 729; DB 33; Length 208577;  
Best Local Similarity 89.4%; Pred. No. 2e-115;  
Matches 815; Conservative 0; Mismatches 92; Indels 5; Gaps 3;

QY 1 ggcgtcgaatgactgcgaaccgcgttcgcagagagccttcgttcgttcgagccnccgga 60  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 13818 GGCCTGCTGAGATGATGAGACAGGTTGCAGAGA--ATCTGGCTTGACCCACACAGGA 13874

QY 61 agaactacgagattgcccagaagactacctctgagtgaagtgaaagtgaggagcaaa 120  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 13875 AGAAGTACTGATTTGGCCAAAAGTACCTTGTAGTAAATGAAATGAGGGGAGCAG 13934

QY 121 tgatggagagaaagcaatcnaagcttcctggaagcaatcagttcccttcgtgaaagaa 180  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 13935 TGATGGAAGAGAAAGCAATCAAAAGCTTGTGAAGCATCTTCCCTGATGAGAAAGAA 13994

QY 181 taaggcgaatgctgcgnaagctgcgagctgagctgagctgagctgagctgagctgag 240  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 13995 TAGGGGAATTTGGCTGAGAGCTGAGGCTAGCTGAGTGAAGTGAAGTTCAGGTTCAG 14054

QY 241 ttctgaagatcagagaaagctgcttccttcgagatcgttcgttcgttcgttcgttc 300  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14055 TTCTGAAGATCAGAGAAAGACGCGGCTTCAGATGCTTGAGCCCGTTAAACTTC 14114

QY 301 atcttccttgccacgttgaaaagcaactgagttagatcnaatcaagaagacagtga 360  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14115 ATCTCTTTGGCCACTGTAAAGCAACTGAATGAGTCAATCAAAAGAGGTGTGGA 14174

QY 361 gttcccttcgaaagagagatgagatccacagagaa-tagcatcattaaaa- 419  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14175 GTTACTCTTTAAAGAAAAATTGAACAGATCCACAGAGAAATACCATTAAC 14234

QY 419 cgcacaagctcctcccaaatggaacctgtctcctgaagaccggaagcagcagct 478  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14235 CTCACAGTCTCTCCAAATGGACCCATCATCTCTGAAGAACACGACGACAGCT 14294

QY 479 ggtttcccttgagaagagagcagcagccatgctcccatgtaacatgctcagtg 538  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14295 GGTTTTCCCTCGGGAAGGACGACGACCATTCCTCCATGAATGCGCTCACTG 14354

QY 539 ctggaagcgaagaaactccctcgagcaggaatttcaacctccatcaagaagca 598  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14355 CTGGAAGGCAAGAACTCCCTCGAGAGGAATTTTAACCTCTCCATTAAGAACAGCA 14414

QY 599 gccagtagcagacccttactgaccctctgtaaaagcctctctcgaagcagcagct 658  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14415 GCCAGTAGCAGATCTTACTGACTCCCATGGAAGAAAGCCCTCTCCAGGCAAGCAGCT 14474

QY 659 agaagagcgaagtgcgacgagcagagctcagaggctcgtggctctgcagtcacta 718  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14475 GGAAGAGCAAGATGACCGACGAGAGCTTCAAGAGGCTCGGCTCTGCAGTCTACTA 14534

QY 719 tgaagcagagctcgaagagagaaatcnaaagttaaaagtatacaaaagtcgagaa 778  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 14535 TGAAGCCAGAGCTGGAAGAGGAATCAAAAGTAAAGTATCAAAAGTCGTGAA 14594

QY 779 gaaaggaagcgaagaaagcccttaaaagagttgagcagctggaagagttatccagc 838



Db	14595	GAAAGAAAGGCGCAAGAAAGCCTTAAAGAGTTTGAGCAGCTACAGAAAGGTAAATCCAAAC	14654
Qy	839	tgcgcactagaagaacgaagaaggaaggaaggaaggaaggaaggaaggaaggaacaa	898
Db	14655	TGTGGCACTGTGGAAGAAATGGAAAAATTGCAAAATGCGCAAGATGCAAAAGATGAGCCT	14714
Qy	899	ggagaagaagaa	910
Db	14715	TAAGCACCAAAA	14726
RESULT	4		
LOCUS	G22194/c		
DEFINITION	G22194	472 bp	DNA
ACCESSION	G22194	human STS	WT-13764, sequence tagged site.
VERSION	G22194.1	GI:1342520	
KEYWORDS	STS; STS sequence; primer; sequence tagged site.		
SOURCE	human STS derived from sequences in dbEST and the Unigene collection.		
ORGANISM	Homo sapiens.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 472)		
TITLE	Hudson, T.		
JOURNAL	Whitehead Institute/MIT Center for Genome Research; Physically Mapped STS		
COMMENT	unpublished (1995)		
	Contact: Thomas Hudson		
	Whitehead Institute/MIT Center for Genome Research		
	Whitehead Institute for Biomedical Research		
	9 Cambridge Center, Cambridge MA 02142 USA		
	Tel: 617 252 1900		
	Fax: 617 252 1902		
	Email: thudson@genome.wi.mit.edu		
	Primer A: CAATGCCACAGTTGGATTAA		
	Primer B: CGGCTCTGCACTCTACTA		
	STS size: 150		
	PCR Profile:		
	Presoak:		
	Denaturation:		
	Annealing: 56 degrees C		
	Polymerization:		
	PCR Cycles: 35		
	Thermal Cycler:		
	Template: 10 ng		
	Primer: each 5 pM		
	dNTPs: each 4 mM		
	Tag Polymerase: 0.025 units/ul		
	Total Vol: 20 ul		
	Buffer:		
	MgCl2: 1.5 mM		
	KCl: 50 mM		
	Tris-HCL: 10 mM		
	pH: 9.3		
FEATURES	Derived from dbEST (genbank accession R54592).		
SOURCE	Location/Qualifiers		
	1..472		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/map="145.4 CR from top of Chr13 linkage group"		
	20..169		
	primer_bind		
	complement(150..169)		
	83 a		
	111 c		
	118 g		
	157 t		
	3 others		
BASE COUNT			
ORIGIN			

Query Match	23.8%	Score 308.6	DB 13	Length 472
Best Local Similarity	86.7%	Pred. No. 1.1e-43		
Matches 385	Conservative 0	Mismatches 52	Indels 7	Gaps 4

QY	429	ctctcnaatgagaccctgtcgtccctgaagaaccgcagcagcagcagctgtttt--c	486
DB	446	CTCCCAATTTGGGAGACCCCTATCATCTGTGAAGAACGACGAGCAGACAGNCTGTTTTCC	387
QY	487	cccttgagaaagagagccca--t-tgtcccatgtgaacatgtg-ctcagtgcttg	542
DB	386	CCCTTGGGAGAGGACGACGCCACCATTTGCTCCCATTTGAACATCGCCTCAGTGGCTGG	327
QY	543	aaggaagaac-ccccctggagcagaatatttcaactccctccaaataaacaacc	601
DB	326	AAGGCAAGACTCCCCCTGGAGCAGGAATTTTAACTCCCTCCTAATGAACAACAGCAGC	267
QY	602	agtgcagaccccttacttgcacccctgttgaaaggcctctctccagccatgactaga	661
DB	266	AGTGACAGATCCTTTACTGACTCCCATGGAAGGCGCTCTCCCAAGCATGAGCTGGA	207
QY	662	agaagcacaagatgagcagcagcagcttccaaggctcgggctcgtcagctcactatga	721
DB	206	AGAGCAAAAGATGACCCGAGCAGACTTCAGAGGCGCTCGGCGCTGCACTCACTAAGA	147
QY	722	ngccaagctcgaagaagaagaatcmaagttaaagrtacaaagtctgaaga	781
DB	146	GGCCAAAGCTCCAAAGAGACAAATAATCAAAAGTAAAGTATCAAAAGTGTAAAGAA	87
QY	782	aggaagagccaaagaagccctaaagaagtltagcagctgcggaaggttaatccagctgc	841
DB	86	AGGAAGGCGCAAGAAAGCCTTAAAGAGTTGAGCAGCTACAGAAAGTTAATCCAACTGT	27
QY	842	cgcactagaagaacgaagaaaga	865
DB	26	GGCATTGGAGAAATGAAAAAAA	3

RESULT	5
LOCUS	166494/c
DEFINITION	Sequence 14 from patent US 5670367.
ACCESSION	166494
VERSION	166494.1
KEYWORDS	GI:2724471
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 7218)
TITLE	Donner,F., Scheiflinger,F. and Falkner,F.Gunter.
JOURNAL	Recombinant fowlpox virus
FEATURES	Patent: US 5670367-A 14-23-SEP-1997;
	Location/Qualifiers
	1..7218
BASE COUNT	1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN	/organism="unknown"

Query Match	16.6%	Score 215.4	DB 5	Length 7218
Best Local Similarity	5.1%	Pred. No. 5.6e-28		
Matches 21	Conservative 354	Mismatches 33	Indels 0	Gaps 0

QY	846	ctagaagaacgaagaagagaggaaggaagggagagagagaagaagaagacagaagaag	905
DB	1466	CAAGTAGTTAAAGACATAGAAATTTGCTACRRRRRRRRRRRRRRRRRRRRRRRRRRRR	1407
QY	906	aagaagaagaagaggaagagaagaaagaagaagaagaagaagaagaagaagaagaa	965
DB	1406	RR	1347
QY	966	agaaaaggaagaaagaaagaaaggaagaagaagaagaactaagaagaagagag	1025

[illegible]

5138	5147:	gap of unknown length
5148	5640:	contig of 493 bp in length
5641	5650:	gap of unknown length
5651	6146:	contig of 496 bp in length
6147	6156:	gap of unknown length
6157	6649:	contig of 493 bp in length
6650	6659:	gap of unknown length
6660	7154:	contig of 495 bp in length
7155	7164:	gap of unknown length
7165	7658:	contig of 494 bp in length
7659	7668:	gap of unknown length
7669	8031:	contig of 363 bp in length
8032	8041:	gap of unknown length
8043	8533:	contig of 492 bp in length
8534	8543:	gap of unknown length
8544	8861:	contig of 318 bp in length
8862	8871:	gap of unknown length
8872	9365:	contig of 492 bp in length
9364	9375:	gap of unknown length
9374	9638:	contig of 265 bp in length
9639	9648:	gap of unknown length
9649	10141:	contig of 493 bp in length
10142	10151:	gap of unknown length
10152	10645:	contig of 492 bp in length
10644	10653:	gap of unknown length
10654	11145:	contig of 492 bp in length
11146	11155:	gap of unknown length
11156	11647:	contig of 492 bp in length
11648	11657:	gap of unknown length
11658	12151:	contig of 494 bp in length
12152	12161:	gap of unknown length
12162	12579:	contig of 418 bp in length
12580	12589:	gap of unknown length
12590	13083:	contig of 494 bp in length
13084	13093:	gap of unknown length
13094	13565:	contig of 472 bp in length
13566	13575:	gap of unknown length
13576	14067:	contig of 492 bp in length
14068	14077:	gap of unknown length
14078	14570:	contig of 493 bp in length
14571	14580:	gap of unknown length
14581	14952:	contig of 372 bp in length
14953	14962:	gap of unknown length
14963	15455:	contig of 493 bp in length
15456	15465:	gap of unknown length
15466	15586:	contig of 121 bp in length
15587	15596:	gap of unknown length
15597	16090:	contig of 494 bp in length
16091	16100:	gap of unknown length
16101	16592:	contig of 492 bp in length
16593	16602:	gap of unknown length
16603	16781:	contig of 179 bp in length
16782	16791:	gap of unknown length
16792	17285:	contig of 494 bp in length
17286	17295:	gap of unknown length
17296	17793:	contig of 498 bp in length
17794	17803:	gap of unknown length
17804	18296:	contig of 493 bp in length
18297	18306:	gap of unknown length
18307	18800:	contig of 494 bp in length
18801	18810:	gap of unknown length
18811	19303:	contig of 493 bp in length
19304	19313:	gap of unknown length
19314	19807:	contig of 494 bp in length
19808	19817:	gap of unknown length
19818	19915:	contig of 98 bp in length
19916	19925:	gap of unknown length
19926	20191:	contig of 266 bp in length
20192	20201:	gap of unknown length
20202	20227:	contig of 26 bp in length
20228	20237:	gap of unknown length
20238	20518:	contig of 281 bp in length
20519	20528:	gap of unknown length

```
* 20529 21208: contig of 680 bp in length
* 21209 21218: gap of unknown length
* 21219 21712: contig of 494 bp in length
* 21713 21722: gap of unknown length
* 21723 22218: contig of 496 bp in length
* 22219 22228: gap of unknown length
* 22229 22722: contig of 494 bp in length
* 22723 22992: gap of unknown length
* 22733 23002: contig of 260 bp in length
* 22993 23446: gap of unknown length
* 23003 23446: contig of 444 bp in length
* 23447 23456: gap of unknown length
* 23457 23949: contig of 493 bp in length
* 23950 23959: gap of unknown length
* 23960 24452: contig of 493 bp in length
* 24453 24462: gap of unknown length
* 24463 24956: contig of 494 bp in length
* 24957 24966: gap of unknown length
* 24967 25458: contig of 492 bp in length
* 25459 25468: gap of unknown length
* 25469 25960: contig of 492 bp in length
* 25970 25970: gap of unknown length
* 25971 26463: contig of 493 bp in length
* 26464 26473: gap of unknown length
* 26474 26967: contig of 494 bp in length
* 26968 26977: gap of unknown length
* 27472 27472: contig of 495 bp in length
* 27473 27482: gap of unknown length
* 27483 27975: contig of 493 bp in length
* 27976 27985: gap of unknown length
* 27986 28362: contig of 377 bp in length
* 28363 28372: gap of unknown length
* 28373 28492: contig of 120 bp in length
* 28493 28502: gap of unknown length
* 28503 28994: contig of 492 bp in length
* 28995 29004: gap of unknown length
* 29005 29678: contig of 674 bp in length
* 29679 29688: gap of unknown length
* 29689 30073: contig of 385 bp in length
* 30074 30083: gap of unknown length
* 30084 30576: contig of 493 bp in length
* 30577 30586: gap of unknown length
* 30587 31078: contig of 492 bp in length
* 31079 31088: gap of unknown length
* 31089 31583: contig of 495 bp in length
* 31584 31593: gap of unknown length
* 31594 32088: contig of 495 bp in length
* 32089 32098: gap of unknown length
* 32099 32593: contig of 495 bp in length
* 32594 32603: gap of unknown length
* 32604 33098: contig of 495 bp in length
* 33099 33108: gap of unknown length
* 33109 33604: contig of 496 bp in length
* 33605 33614: gap of unknown length
* 33615 34111: contig of 497 bp in length
* 34112 34121: gap of unknown length
* 34122 34614: contig of 493 bp in length
* 34615 34624: gap of unknown length
* 34625 35118: contig of 494 bp in length
* 35119 35128: gap of unknown length
* 35129 35619: contig of 491 bp in length
* 35620 35629: gap of unknown length
* 35630 36122: contig of 493 bp in length
* 36123 36132: gap of unknown length
* 36133 36561: contig of 429 bp in length
* 36562 36571: gap of unknown length
* 36572 37062: contig of 491 bp in length
* 37063 37072: gap of unknown length
* 37073 37567: contig of 495 bp in length
* 37568 37577: gap of unknown length
* 37578 38072: contig of 495 bp in length
* 38073 38082: gap of unknown length
* 38083 38576: contig of 494 bp in length
```

```
* 38577 38586: gap of unknown length
* 38587 38883: contig of 297 bp in length

Query Match 14.7%; Score 190.8; DB 52; Length 131025;
Best Local Similarity 59.6%; Pred. No. 6.1e-24;
Matches 336; Conservative 0; Mismatches 226; Indels 2; Gaps 1;

QY 735 agagagaagaatacnaaagttaaagttacacaaagtcgtgaagaagaagccaaag 794
Db 129823 AAAAAAAAAAGAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 129764

QY 795 aaagccctaaagatttgacagctgcggaagttatccagctgcgactagaagaa 854
Db 129763 AAAAAAAAAAATATTAACCCCCCCCAAGAGAGAAAAAAGAAAAAGAAAAAAG 129704

QY 855 cgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 914
Db 129703 AAGAAAAAGGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 129644

QY 915 gaaggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 974
Db 129643 AGAAAGCGAAAAAAGAGAAAAAAGAAAAAGAAAGAAA--AAGAAAAAGAAAGAAAAAGA 129586

QY 975 gaaggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaataag 1034
Db 129585 AAAAAAGAAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAAGAAAAAGAAAG 129526

QY 1035 aaaggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1094
Db 129525 AAAAAAGGAGAAAAACAAAAAGAAAAATGAGAGAGAAAAAGAAAAAAGAAAGAA 129466

QY 1095 agacthagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1154
Db 129465 GAAAAAGAAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAAGAAAAAGCGC 129406

QY 1155 agaagaagaataaagaagaagaagaagaagaagaagaagaagaagaagaaga 1214
Db 129405 AAAAAAAGAAAAAAGAAAAAGAAAAAAGAAAAAAGAAAAAGAAAAAGAAAAAAG 129346

QY 1215 aaaggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1274
Db 129345 AAAAAAGAGAAAAAAGAAAAAGAAAAAAGAAAAAAGAAAAAGAAAAAAGAAAGC 129286

QY 1275 ggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1298
Db 129285 GAAAAAGAAAAAAGAAAAAGAAAAAGAAAA 129262

RESULT 7
LOCUS AC021195/c 53156 bp DNA HTG 14-JAN-2000
DEFINITION Homo sapiens clone Rp11-780K2, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC021195
VERSION AC021195.1 GI:6693387
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 53156)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 53156)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

COMMENT
* Center project name: NH0563P16.
* NOTE: This record contains 94 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
```

\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 424: contig of 424 bp in length  
\* 425 434: gap of unknown length  
\* 435 827: contig of 393 bp in length  
\* 828 837: gap of unknown length  
\* 838 1335: contig of 498 bp in length  
\* 1336 1345: gap of unknown length  
\* 1346 1858: contig of 513 bp in length  
\* 1859 1868: gap of unknown length  
\* 1869 2196: contig of 328 bp in length  
\* 2197 2206: gap of unknown length  
\* 2207 2606: contig of 400 bp in length  
\* 2607 2616: gap of unknown length  
\* 2617 3130: contig of 514 bp in length  
\* 3131 3140: gap of unknown length  
\* 3141 3396: contig of 286 bp in length  
\* 3397 3406: gap of unknown length  
\* 3407 3921: contig of 515 bp in length  
\* 3922 3931: gap of unknown length  
\* 3932 4446: contig of 515 bp in length  
\* 4447 4456: gap of unknown length  
\* 4457 4972: contig of 516 bp in length  
\* 4973 4982: gap of unknown length  
\* 4983 5633: contig of 651 bp in length  
\* 5634 5643: gap of unknown length  
\* 5644 6339: contig of 696 bp in length  
\* 6340 6349: gap of unknown length  
\* 6350 6864: contig of 515 bp in length  
\* 6865 6874: gap of unknown length  
\* 6875 7388: contig of 514 bp in length  
\* 7389 7398: gap of unknown length  
\* 7399 7912: contig of 514 bp in length  
\* 7913 7922: gap of unknown length  
\* 7923 8436: contig of 514 bp in length  
\* 8437 8446: gap of unknown length  
\* 8447 9087: contig of 641 bp in length  
\* 9088 9097: gap of unknown length  
\* 9098 9738: contig of 641 bp in length  
\* 9739 9748: gap of unknown length  
\* 9749 10262: contig of 514 bp in length  
\* 10263 10272: gap of unknown length  
\* 10273 10934: contig of 662 bp in length  
\* 10935 10944: gap of unknown length  
\* 10945 11577: contig of 633 bp in length  
\* 11578 11587: gap of unknown length  
\* 11588 12284: contig of 697 bp in length  
\* 12285 12294: gap of unknown length  
\* 12295 12983: contig of 689 bp in length  
\* 12984 12993: gap of unknown length  
\* 12994 13602: contig of 609 bp in length  
\* 13603 13612: gap of unknown length  
\* 13613 14240: contig of 628 bp in length  
\* 14241 14250: gap of unknown length  
\* 14251 14941: contig of 691 bp in length  
\* 14942 14951: gap of unknown length  
\* 14952 15648: contig of 697 bp in length  
\* 15649 15658: gap of unknown length  
\* 15659 16172: contig of 514 bp in length  
\* 16173 16182: gap of unknown length  
\* 16183 16696: contig of 514 bp in length  
\* 16697 16706: gap of unknown length  
\* 16707 17155: contig of 449 bp in length  
\* 17156 17165: gap of unknown length  
\* 17166 17597: contig of 432 bp in length  
\* 17598 17607: gap of unknown length  
\* 17608 18053: contig of 446 bp in length

\* 18054 18063: gap of unknown length  
\* 18064 18885: contig of 822 bp in length  
\* 18886 18895: gap of unknown length  
\* 18896 19597: contig of 702 bp in length  
\* 19598 19607: gap of unknown length  
\* 19608 20511: contig of 904 bp in length  
\* 20512 20521: gap of unknown length  
\* 20522 21216: contig of 695 bp in length  
\* 21217 21226: gap of unknown length  
\* 21227 21676 21685: gap of unknown length  
\* 21686 22359: contig of 674 bp in length  
\* 22360 22369: gap of unknown length  
\* 22370 22882: contig of 513 bp in length  
\* 22883 22892: gap of unknown length  
\* 22893 23381: contig of 489 bp in length  
\* 23382 23391: gap of unknown length  
\* 23392 24315: contig of 924 bp in length  
\* 24316 24325: gap of unknown length  
\* 24326 24839: contig of 514 bp in length  
\* 24840 24849: gap of unknown length  
\* 24850 25364: contig of 515 bp in length  
\* 25365 25374: gap of unknown length  
\* 25375 26204: contig of 830 bp in length  
\* 26205 26214: gap of unknown length  
\* 26215 26976: contig of 762 bp in length  
\* 26977 26986: gap of unknown length  
\* 26987 27671: contig of 685 bp in length  
\* 27672 27681: gap of unknown length  
\* 27682 28585: contig of 904 bp in length  
\* 28586 28595: gap of unknown length  
\* 28596 29262: contig of 667 bp in length  
\* 29263 29272: gap of unknown length  
\* 29273 29947: contig of 675 bp in length  
\* 29948 29957: gap of unknown length  
\* 29958 30471: contig of 514 bp in length  
\* 30472 30481: gap of unknown length  
\* 30482 31118: contig of 637 bp in length  
\* 31119 31128: gap of unknown length  
\* 31129 31642: contig of 514 bp in length  
\* 31643 31652: gap of unknown length  
\* 31653 32167: contig of 515 bp in length  
\* 32168 32177: gap of unknown length  
\* 32178 33388: contig of 211 bp in length  
\* 32389 33398: gap of unknown length  
\* 32399 33052: contig of 654 bp in length  
\* 33053 33062: gap of unknown length  
\* 33063 33604: contig of 542 bp in length  
\* 33605 33614: gap of unknown length  
\* 33615 34128: contig of 514 bp in length  
\* 34129 34138: gap of unknown length  
\* 34139 34650: contig of 512 bp in length  
\* 34651 34660: gap of unknown length  
\* 34661 35174: contig of 514 bp in length  
\* 35175 35184: gap of unknown length  
\* 35185 35698: contig of 514 bp in length  
\* 35699 35708: gap of unknown length  
\* 35709 36222: contig of 514 bp in length  
\* 36233 36232: gap of unknown length  
\* 36233 36746: contig of 514 bp in length  
\* 36747 36756: gap of unknown length  
\* 36757 37270: contig of 514 bp in length  
\* 37271 37280: gap of unknown length  
\* 37281 37771: contig of 491 bp in length  
\* 37772 37781: gap of unknown length  
\* 37782 38285: contig of 504 bp in length  
\* 37826 38295: gap of unknown length  
\* 38286 38810: contig of 515 bp in length  
\* 38811 38820: gap of unknown length  
\* 38821 39335: contig of 515 bp in length  
\* 39336 39345: gap of unknown length  
\* 39346 39807: contig of 462 bp in length  
\* 39808 39817: gap of unknown length

```

*      39818      40332: contig of 515 bp in length
*      40333      40342: gap of unknown length
*      40343      40386: contig of 644 bp in length
*      40987      40996: gap of unknown length
*      40997      41510: contig of 514 bp in length
*      41511      41520: gap of unknown length
*      41521      42034: contig of 514 bp in length
*      42035      42044: gap of unknown length
*      42045      42558: contig of 514 bp in length
*      42559      42569: gap of unknown length
*      42569      43082: contig of 514 bp in length
*      43083      43092: gap of unknown length
*      43093      43607: contig of 515 bp in length
*      43608      43617: gap of unknown length
*      43618      44132: contig of 515 bp in length
*      44133      44142: gap of unknown length
*      44143      44657: contig of 515 bp in length
*      44658      44667: gap of unknown length
*      44668      45079: contig of 412 bp in length
*      45080      45089: gap of unknown length
*      45090      45786: contig of 697 bp in length
*      45787      45796: gap of unknown length
*      45797      46311: contig of 515 bp in length
*      46312      46321: gap of unknown length
*      46322      46835: contig of 514 bp in length
*      46836      46845: gap of unknown length
*      46846      47359: contig of 514 bp in length
*      47360      47369: gap of unknown length
*      47370      47841: contig of 472 bp in length
*      47842      47851: gap of unknown length
*      47852      48365: contig of 514 bp in length

```

Query Match 14.6%; Score 189.6; DB 44; Length 53156;  
 Best Local Similarity 56.5%; Pred. No. 1.le-23;

Matches 336; Conservative 0; Mismatches 259; Indels 0; Gaps 0;

```

QY 704 tctgcagtcctctctgtgagcgaagctcgaagagagagaatcnaagtaagta 763
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21258 TTTGGCCTATTTCGTCCTCCATCTGGGAAGACNNNNNNNNNNAAAAA 21199
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 764 tcacaaagtcgtaagaaaggaagcacaagaagccctaagaagtttgagcagctgcg 823
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21198 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGA 21139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 824 gaaggttaatccagctgcgcgactagaagacgaagaagaaggaagagagag 883
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21138 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 21079
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 884 aaagaagaagaagaagaagaagaagaagaagagagagaagaagaagaaga 943
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21078 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 21019
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 944 agaggaagaaggaaggaagaagaagaagaagaagaagaagaagaagaag 1003
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21018 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 20959
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1004 aagaactaagaaggaaggaagaagaagaagaagaagaagaagaagaaga 1063
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20958 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 20899
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1064 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1123
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20898 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 20839
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1124 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20838 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 20779
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1184 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 20778 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 20719
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1244 aaagcggaagaagaagaagaagaagaagaagaagaagaagaagaaga 1298

```

Db 20718 AA 20664

RESULT 8  
 AC013349/c  
 LOCUS  
 DEFINITION  
 AC013349  
 ACCESSION  
 AC013349.2 GI:6910730  
 KEYWORDS  
 HTG; HTGS\_PHASE0.  
 SOURCE  
 ORGANISM  
 Homo sapiens

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 2 (bases 1 to 129404)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
 Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,  
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collumore,A.,  
 Cooke,P., Deatellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
 Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,  
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Lechickzy,J., Lien,C., Locke,K., Macdonald,P., Marquis,N.,  
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,  
 Morrow,J., Maylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
 Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
 Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
 Wymad,D., Ye,W.J., Zimmerman,A. and Zody,M.

Direct Submission  
 Submitted (06-NOV-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Feb 6, 2000 this sequence version replaced gi:672406.  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 Project Information  
 Center project name: LA134  
 Center clone name: 22\_K.1

NOTE: This record contains 151 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.  
 \* 1  
 \* 931 930: contig of 930 bp in length  
 \* 1856 1855: gap of unknown length  
 \* 2759 2759: gap of 925 bp in length  
 \* 3706 3706: gap of unknown length  
 \* 4566 4566: gap of 904 bp in length  
 \* 5444 5444: gap of unknown length  
 \* 5457 5457: gap of 947 bp in length  
 \* 5457 5456: gap of unknown length  
 \* 5457 5456: contig of 860 bp in length  
 \* 5457 5457: gap of unknown length  
 \* 5457 5457: contig of 878 bp in length  
 \* 5457 5457: gap of unknown length



[illegible]

RESULT	9			
LOCUS	AC010649/c			
DEFINITION	AC010649	198664 bp	DNA	HTG
ACCESSION	AC010649			14-JAN-2000
VERSION	AC010649.2	GI:6693147		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
AUTHORS	1 (bases 1 to 198664)			
TITLE	DOE Joint Genome Institute.			
JOURNAL	Sequencing of Human Chromosome 19			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 198664)			
TITLE	DOE Joint Genome Institute.			
JOURNAL	Direct Submission			
COMMENT	Submitted (16-SEP-1999) Production Sequencing Facility. DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
	On Jan 14, 2000 this sequence version replaced gi:5902296.			

```

ON Jan 14, 2000 this sequence version replaced g1:3502256.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Consensus quality: 111206 bases at least Q40
Consensus quality: 137820 bases at least Q30
Consensus quality: 148360 bases at least Q20
Estimated insert size: 198664; sum-of-contigs estimation
Estimated insert size: 38000; agarose-fp estimation
Quality coverage: 25.37% in Q20 bases; sum-of-contigs estimation
Quality coverage: 4.85% in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 75 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

* This record will be updated with the finished sequence		
* as soon as it is available and the accession number will		
* be preserved.		
1	1323:	contig of 1323 bp in length
*		gap of unknown length
*	1324	2429: contig of 1106 bp in length
*		gap of unknown length
*	2430	3618: contig of 1189 bp in length
*		gap of unknown length
*	3619	4982: contig of 1364 bp in length
*		gap of unknown length
*	4983	6346: contig of 1364 bp in length
*		gap of unknown length
*	6347	7470: contig of 1124 bp in length
*		gap of unknown length
*	7471	8756: contig of 1286 bp in length
*		gap of unknown length
*	8757	10044: contig of 1288 bp in length
*		gap of unknown length
*	10045	11650: contig of 1606 bp in length
*		gap of unknown length
*	11651	13132: contig of 1482 bp in length
*		gap of unknown length
*	13133	14366: contig of 1234 bp in length
*		gap of unknown length
*	14367	15620: contig of 1254 bp in length
*		gap of unknown length
*	15621	17023: contig of 1403 bp in length
*		gap of unknown length
*	17024	18609: contig of 1586 bp in length
*		gap of unknown length
*	18610	19741: contig of 1132 bp in length
*		gap of unknown length
*	19742	22125: contig of 2384 bp in length
*		gap of unknown length
*	22126	23229: contig of 1104 bp in length
*		gap of unknown length
*	23230	24473: contig of 1244 bp in length
*		gap of unknown length
*	24474	25554: contig of 1181 bp in length
*		gap of unknown length
*	25555	26756: contig of 1102 bp in length
*		gap of unknown length
*	26757	28226: contig of 1470 bp in length
*		gap of unknown length
*	28227	29719: contig of 1493 bp in length
*		gap of unknown length
*	29720	31090: contig of 1371 bp in length
*		gap of unknown length
*	31091	32423: contig of 1333 bp in length
*		gap of unknown length
*	32424	34191: contig of 1768 bp in length
*		gap of unknown length
*	34192	35479: contig of 1288 bp in length
*		gap of unknown length
*	35480	36788: contig of 1309 bp in length
*		gap of unknown length
*	36789	38007: contig of 1219 bp in length
*		gap of unknown length
*	38008	39215: contig of 1208 bp in length
*		gap of unknown length
*	39216	40918: contig of 1703 bp in length
*		gap of unknown length
*	40919	42203: contig of 1285 bp in length
*		gap of unknown length
*	42204	43550: contig of 1347 bp in length
*		gap of unknown length
*	43551	44654: contig of 1104 bp in length
*		gap of unknown length
*	44655	46253: contig of 1599 bp in length
*		gap of unknown length
*	46254	48291: contig of 2038 bp in length
*		gap of unknown length

gap of unknown length

48232	49705:	contig of 1414 bp in length
49706	50985:	gap of unknown length
50986	52437:	contig of 1280 bp in length
52438	54118:	gap of unknown length
54119	55962:	contig of 1442 bp in length
55963	57087:	gap of unknown length
57088	58591:	contig of 1125 bp in length
58592	60625:	gap of unknown length
60626	61850:	contig of 1225 bp in length
61851	63794:	gap of unknown length
63795	65350:	contig of 1944 bp in length
65351	67281:	gap of unknown length
67282	68471:	contig of 1556 bp in length
68472	70100:	gap of unknown length
70101	72256:	contig of 1639 bp in length
72257	74188:	gap of unknown length
74189	76758:	contig of 1932 bp in length
76759	78733:	gap of unknown length
78734	80979:	contig of 2570 bp in length
80980	82507:	gap of unknown length
82508	85571:	contig of 1528 bp in length
85572	87481:	gap of unknown length
87482	89902:	contig of 3064 bp in length
89903	92467:	gap of unknown length
92468	96571:	contig of 2565 bp in length
96572	98660:	gap of unknown length
98661	101042:	contig of 2421 bp in length
101043	103468:	gap of unknown length
103469	105074:	contig of 2426 bp in length
105075	111367:	gap of unknown length
111368	115229:	contig of 1606 bp in length
115230	119240:	gap of unknown length
119241	124263:	contig of 4011 bp in length
124264	127866:	gap of unknown length
127867	136874:	contig of 5023 bp in length
136875	141030:	gap of unknown length
141031	158945:	contig of 4156 bp in length
		gap of unknown length
		contig of 17915 bp in length

FEATURES	Source	Location/Qualifiers
* * *	158946	gap of unknown length
* * *	166285	contig of 7340 bp in length
* * *	166286	gap of unknown length
* * *	177898	contig of 11613 bp in length
* * *	177899	gap of unknown length
* * *	198664	contig of 20766 bp in length.
FEATURES		
Source		
BASE COUNT	52403 a 45760 c 45438 g 53746 t 1317 others	
ORIGIN		
Query Match	14.1%	Score 182.4 : DB 44; Length 198664;
Best Local Similarity	57.4%	Pred. No. 1.6e-22;
Matches 321: Conservative	0; Mismatches 238; Indels 0; Gaps	
Qy	737	aagaaagaaatcnaaagttaaaglatcacaaagtcgtgaagaaagaaagcccaagaa 796
Db	38647	AGAGAAAGGAGCAGCAGAGCAAAAGAGAGAGCAAGAAAAGCAGAAAGAAAAGGGGAG 385888
Qy	797	agccctaaagaagtttgagcagctgcgcgaaggttaatccaagctgcgcgaactagaagaacg 856
Db	38587	AGAGGAGAGAGAGAGAGGAGGAGGAGAGAGAAAAGAGCGGAGAGAGAAAAGGGA 385288
Qy	857	aagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 916
Db	38527	AAAGGAGAGGAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 38468
Qy	917	aaggaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 976
Db	38467	AAAGGAGAGGAG 38408
Qy	977	aggaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 1036
Db	38407	AAAAAG 38348
Qy	1037	ggaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaag 1096
Db	38347	AAAAAG 38288
Qy	1097	aactgaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaag 1156
Db	38287	AAAAAG 38228
Qy	1157	aagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 1216
Db	38227	AAAAAG 38168
Qy	1217	aggaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaag 1276
Db	38167	AAAAAG 38108
Qy	1277	aagaagaaagaaagaaagaa 1295
Db	38107	AAAAAG 38089
RESULT 10		
AC009524		
LOCUS	AC009524	155284 bp DNA HTG 26-AUG-1999
DEFINITION	AC009524	Human sapiens chromosome 15 clone BAC 31701 map 15q24, LOW-PASS
SEQUENCE	AC009524	SEQUENCE SAMPLING.
VERSION	AC009524.1	GI:5776572
KEYWORDS	HTG: HTGS_PHSO.	
SOURCE	human.	
ORGANISM	Human sapiens	
	Eutheria; Primates; Chordata; Craniata; Vertebrata; Mammalia;	
	Eutheria; Primates; Catarrhini; Hominoidea; Homo.	



REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 155284)	Rowen, L., Madan, A., Olin, S., Abbasi, N., Baradarani, L., Birditt, B., Bloom, S., Dors, M., Dickhoff, R., Harrison, G., James, R., Lasky, S., Madan, A., Ratcliffe, A., Shafter, T. and Hood, L.	Sequencing of human chromosome 15 D15S160-D15S115 region	Unpublished	2 (bases 1 to 155284)	Rowen, L., Madan, A., Olin, S., Abbasi, N., Baradarani, L., Birditt, B., Bloom, S., Dors, M., Dickhoff, R., Harrison, G., James, R., Lasky, S., Madan, A., Ratcliffe, A., Shafter, T. and Hood, L.	Sequencing of human chromosome 15 D15S160-D15S115 region	Unpublished
* NOTE: This record contains 192 individual							
* sequencing reads that have not been assembled into							
* contigs. Runs of N are used to separate the reads							
* and the order in which they appear is completely							
* arbitrary. Low-pass sequence sampling is useful for							
* identifying clones that may be gene-rich and allows							
* overlap relationships among clones to be deduced.							
* However, it should not be assumed that this clone							
* will be sequenced to completion. In the event that							
* the record is updated, the accession number will							
* be preserved.							
1	794:	contig of 794 bp in length	gap of unknown length	20155	20968:	contig of 814 bp in length	gap of unknown length
*	795	1602: contig of 808 bp in length	gap of unknown length	*	20969	21768: contig of 800 bp in length	gap of unknown length
*	1603	2437: contig of 835 bp in length	gap of unknown length	*	21769	22582: contig of 814 bp in length	gap of unknown length
*	2438	3257: contig of 820 bp in length	gap of unknown length	*	22583	23385: contig of 803 bp in length	gap of unknown length
*	3258	4053: contig of 796 bp in length	gap of unknown length	*	23386	24200: contig of 815 bp in length	gap of unknown length
*	4054	4865: contig of 812 bp in length	gap of unknown length	*	24201	25005: contig of 805 bp in length	gap of unknown length
*	4866	5659: contig of 794 bp in length	gap of unknown length	*	25006	25821: contig of 816 bp in length	gap of unknown length
*	5660	6468: contig of 809 bp in length	gap of unknown length	*	25822	26627: contig of 806 bp in length	gap of unknown length
*	6469	7263: contig of 795 bp in length	gap of unknown length	*	26628	27436: contig of 809 bp in length	gap of unknown length
*	7264	8077: contig of 814 bp in length	gap of unknown length	*	27437	28229: contig of 793 bp in length	gap of unknown length
*	8078	8872: contig of 795 bp in length	gap of unknown length	*	28230	29054: contig of 825 bp in length	gap of unknown length
*	8873	9684: contig of 812 bp in length	gap of unknown length	*	29055	29852: contig of 798 bp in length	gap of unknown length
*	9685	10479: contig of 795 bp in length	gap of unknown length	*	29853	30662: contig of 810 bp in length	gap of unknown length
*	10480	11292: contig of 813 bp in length	gap of unknown length	*	30663	31461: contig of 799 bp in length	gap of unknown length
*	11293	12090: contig of 798 bp in length	gap of unknown length	*	31462	32271: contig of 810 bp in length	gap of unknown length
*	12091	12904: contig of 814 bp in length	gap of unknown length	*	32272	33067: contig of 796 bp in length	gap of unknown length
*	12905	13695: contig of 791 bp in length	gap of unknown length	*	33068	33880: contig of 813 bp in length	gap of unknown length
*	13696	14508: contig of 813 bp in length	gap of unknown length	*	33881	34674: contig of 794 bp in length	gap of unknown length
*	14509	15312: contig of 804 bp in length	gap of unknown length	*	34675	35488: contig of 814 bp in length	gap of unknown length
*	15313	16128: contig of 816 bp in length	gap of unknown length	*	35489	36286: contig of 798 bp in length	gap of unknown length
*	16129	16926: contig of 798 bp in length	gap of unknown length	*	36287	37097: contig of 811 bp in length	gap of unknown length
*	16927	17739: contig of 813 bp in length	gap of unknown length	*	37098	37896: contig of 799 bp in length	gap of unknown length
*	17740	18538: contig of 799 bp in length	gap of unknown length	*	37897	38707: contig of 811 bp in length	gap of unknown length
*	18539	19353: contig of 815 bp in length	gap of unknown length	*	38708	39517: contig of 810 bp in length	gap of unknown length
*	19354	20154: contig of 801 bp in length	gap of unknown length	*	39518	40333: contig of 816 bp in length	gap of unknown length
*				*	40334	41127: contig of 794 bp in length	gap of unknown length
*				*	41128	41941: contig of 814 bp in length	gap of unknown length
*				*	41942	42737: contig of 796 bp in length	gap of unknown length
*				*	42738	43549: contig of 812 bp in length	gap of unknown length
*				*	43550	44347: contig of 798 bp in length	gap of unknown length
*				*	44348	45160: contig of 813 bp in length	gap of unknown length
*				*	45161	45959: contig of 799 bp in length	gap of unknown length
*				*	45960	46773: contig of 814 bp in length	gap of unknown length
*				*	46774	47575: contig of 802 bp in length	gap of unknown length
*				*	47576	48384: contig of 809 bp in length	gap of unknown length
*				*	48385	49187: contig of 803 bp in length	gap of unknown length



[illegible]



[illegible]

RESULT	13	
LOCUS	HSAC000362	
DEFINITION	HSAC000362 40977 bp ms-DNA	PRI
ACCESSION	Human cosmid g136a094, complete sequence.	12-MAR-1997
VERSION	AC000362	
KEYWORDS	AC000362.1 GI:181561	
SOURCE	HTG.	
	human.	

ORGANISM	<i>Homo sapiens</i> Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 40977)
REFERENCE	Idonoto, S.P., Yu, J.J., Wong, G.K.-S., Magnes, C.L., Green, E.D., Green, P., and Olson, M.V. Large-scale WGD Mapping and Sequencing of Human Chromosome 7 Unpublished (1996) 2 (bases 1 to 40977)
TITLE	
JOURNAL	
REFERENCE	

AUTHORS	Magness, C.L.
TITLE	Direct Submission
JOURNAL	Submitted (12-MAR-1997) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
COMMENT	This entry has been annotated with sequence quality

Verification: This sequence has been verified by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

	ECORI		HindIII		NsiI	
	Map	Seq	Map	Seq	Map	Seq
8150..50		8225..00	4342..30	4337..00	4100..36	4094..00
2000..57		2026..00	2090..70	2115..00	966..40	966..00
2612..00		2622..00	1671..50	1671..00	2803..25	2825..00
3470..00		3545..00	3134..29	3147..00	6224..00	6330..00

[illegible]



Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
 Dearellano, J., Dewar, K., Dodge, S., Domino, M., Doyle, M.,  
 Feneator, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
 Klein, J., Landers, T., Largoczek, K., Lehoczek, J., Levine, R.,  
 Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,  
 McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J.,  
 Meneus, L., Mihova, T., Miranda, C., Mlenga, Y., Morrow, J., Naylor, J.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T. M.,  
 Peterson, J. C., Pierre, N., Pisan, C., Pollara, V., Raymond, C.,  
 Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,  
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Subramanian, A., Talamas, J., Testfaye, S., Theodore, J., Tirrell, A.,  
 Travers, M., Triggilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,  
 Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and  
 Zody, M.

Direct Submission  
 Submitted (11-FEB-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WITBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information -----  
 Center project name: L6671  
 Center clone name: 373\_D\_19

-----

\* NOTE: This record contains 78 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

1	774:	contig of 774 bp in length	*
*		gap of unknown length	*
*	775	1606: contig of 832 bp in length	*
*		gap of unknown length	*
*	1607	2431: contig of 825 bp in length	*
*		gap of unknown length	*
*	2432	3241: contig of 810 bp in length	*
*		gap of unknown length	*
*	3242	4001: contig of 760 bp in length	*
*		gap of unknown length	*
*	4002	4712: contig of 711 bp in length	*
*		gap of unknown length	*
*	4713	5493: contig of 781 bp in length	*
*		gap of unknown length	*
*	5494	6263: contig of 770 bp in length	*
*		gap of unknown length	*
*	6264	7063: contig of 800 bp in length	*
*		gap of unknown length	*
*	7064	7876: contig of 813 bp in length	*
*		gap of unknown length	*
*	7877	8658: contig of 782 bp in length	*
*		gap of unknown length	*
*	8659	9477: contig of 819 bp in length	*
*		gap of unknown length	*
*	9478	10297: contig of 820 bp in length	*
*		gap of unknown length	*
*	10298	11101: contig of 804 bp in length	*
*		gap of unknown length	*
*	11102	11851: contig of 750 bp in length	*
*		gap of unknown length	*





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 01:07:30 ; Search time 6491.04 Seconds

(without alignments)  
-324.012 Million cell updates/sec

**Title:** US-09-502-945-5  
**Perfect score:** 2162  
**Sequence:** 1 cctgcccgcgtcgcgtcgc.....tatttcagcttaaaaaa 2162

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

```
Minimum DB seq length: 0
Maximum DB seq length: 1000000
```

Post-processing: Minimum Match 0%

Database :

1:	gb_ba1:	*
2:	gb_ba2:	*
3:	gb_om:	*
4:	gb_ov:	*
5:	gb_pat:	*
6:	gb_ph:	*
7:	gb_ph1:	*
8:	gb_pl2:	*
9:	gb_pl1:	*
10:	gb_pf2:	*
11:	gb_pf3:	*
12:	gb_to:	*
13:	gb_sys:	*
14:	gb_sy:	*
15:	gb_un:	*
16:	gb_v1:	*
17:	em_fun:	*
18:	em_hum1:	*
19:	em_hum2:	*
20:	em_in:	*
21:	em_om:	*
22:	em_oc:	*
23:	em_ov:	*
24:	em_pst:	*
25:	em_ph:	*
26:	em_p1:	*
27:	em_to:	*
28:	em_sys:	*
29:	em_sy:	*
30:	em_un:	*
31:	em_v1:	*
32:	gb_hg1:	*
33:	gb_hb2:	*
34:	gb_in1:	*
35:	gb_in2:	*
36:	em_ba2:	*
37:	em_ba1:	*
38:	em_hum3:	*
39:	em_hum4:	*
40:	gb_p14:	*
41:	gb_hg3:	*
42:	gb_hg4:	*
43:	gb_hg5:	*
44:	gb_hg6:	*

```

45:  gb_htgt7:*
46:  em_htg1:*
47:  em_htg2:*
48:  em_htg3:*
49:  em_hums:*
50:  gb_p13:*
51:  gb_pr5:*
52:  gb_htg8:*
53:  gb_htg9:*
54:  gb_htg10:*
55:  gb_htg11:*
56:  gb_htg12:*
57:  gb_htg13:*
58:  gb_htg14:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description
	1	2162	100.0	2162	40	AF039699	AF039699 Homo sapi	AF039699 Homo sapi
	2	2078	96.1	2236	40	AF039700	AF039700 Homo sapi	AF039700 Homo sapi
	3	1981.4	91.6	2213	9	AB018687	AB018687 Homo sapi	AB018687 Homo sapi
	4	1843.6	85.3	2283	9	AB006955	AB006955 Homo sapi	AB006955 Homo sapi
C	5	437.4	20.2	161190	11	AC005137	AC005137 Human Chr	AC005137 Human Chr
C	6	105	4.9	463	13	G55504	G55504 SHC-10085	G55504 SHC-10085
C	7	75.8	3.5	7218	5	T66494	T66494 Sequence 14	T66494 Sequence 14
C	8	69.6	3.2	180385	40	AC007461	AC007461 Homo sapi	AC007461 Homo sapi
C	9	68.8	3.2	166171	42	AC011767	AC011767 Homo sapi	AC011767 Homo sapi
10	10	65.2	3.0	451	11	MUSIR3E2	M10668 Mouse DNA	M10668 Mouse DNA
11	11	65.2	3.0	2902	12	HSM800885	AL110228 Homo sapi	AL110228 Homo sapi
12	12	64.8	3.0	16945	44	AC019774	AC019774 Drosophila	AC019774 Drosophila
13	13	63.4	2.9	195385	43	AC016818	AC016818 Homo sapi	AC016818 Homo sapi
C	14	63.2	2.9	28559	16	AF148805	AF148805 Kapozi's	AF148805 Kapozi's
15	15	63.2	2.9	230309	57	AC016891	AC016961 Homo sapi	AC016961 Homo sapi
16	16	63	2.9	170387	33	AL139411	AL139411 Homo sapi	AL139411 Homo sapi
17	17	62.8	2.9	1080	12	MUS9211A	D38613 Mouse 921-1-	D38613 Mouse 921-1-
18	18	62.2	2.9	59686	42	AC016204	AC016204 Homo sapi	AC016204 Homo sapi
C	19	62	2.9	16325	42	AC015144	AC015144 Drosophila	AC015144 Drosophila
C	20	62	2.9	65334	56	AC010699	AC010699 Drosophila	AC010699 Drosophila
C	21	62	2.9	106664	12	AF130357	AF130357 Mus muscu	AF130357 Mus muscu
22	22	62	2.9	132106	56	AC010031	AC010031 Drosophila	AC010031 Drosophila
C	23	61.4	2.8	68130	40	AC004961	AC004961 Homo sapi	AC004961 Homo sapi
C	24	61	2.8	460	12	MUSIR3E1	M10286 Mouse DNA	M10286 Mouse DNA
C	25	61	2.8	78467	40	AC008078	AC008078 Homo sapi	AC008078 Homo sapi
C	26	60.6	2.8	170336	32	AL135899	AL135899 Mus muscu	AL135899 Mus muscu
C	27	60.6	2.8	217025	33	AL138895	AL138895 Homo sapi	AL138895 Homo sapi
C	28	60.4	2.8	90832	40	AC003065	AC003065 Homo sapi	AC003065 Homo sapi
C	29	60.4	2.8	97348	12	AF091216	AF091216 Mus muscu	AF091216 Mus muscu
C	30	60.4	2.8	113843	40	AC005033	AC005033 Homo sapi	AC005033 Homo sapi
C	31	60.4	2.8	133661	16	U93872	U93872 Kapozi's sa	U93872 Kapozi's sa
32	32	60.4	2.8	234523	40	AC006544	AC006544 , complete	AC006544 , complete
C	33	60.2	2.8	110000	33	AL137126.2	Continuation (3 OF	Continuation (3 OF
C	34	59.8	2.8	5719	12	MMCAR51	AF051726 Mus muscu	AF051726 Mus muscu
C	35	59.6	2.8	148700	42	AC007492	AC007492 Homo sapi	AC007492 Homo sapi
C	36	59.6	2.8	183570	54	AC007906	AC007906 Homo sapi	AC007906 Homo sapi
C	37	58.8	2.7	631	12	MMY14822	Y14822 Mus muscu	Y14822 Mus muscu
38	38	58.8	2.7	3489	16	KSU52064	U52064 Kapozi's sa	U52064 Kapozi's sa
C	39	58.8	2.7	32207	5	AR065852	AR065852 Sequence	AR065852 Sequence
C	40	58.8	2.7	37918	41	AC011553	AC011553 Homo sapl	AC011553 Homo sapl
C	41	58.8	2.7	137508	16	KSU75698	U75698 Kapozi's sa	U75698 Kapozi's sa
C	42	58.6	2.7	132400	41	AC007042	AC007042 Homo sapi	AC007042 Homo sapi
C	43	58.6	2.7	170021	45	AC016791	AC016791 Mus muscu	AC016791 Mus muscu
C	44	58.4	2.7	170095	9	HSEGRF4G	Y13901 Homo saplen	Y13901 Homo saplen
C	45	58.4	2.7	118593	11	HS2886L1	Z82196 Human DNA s	Z82196 Human DNA s

RESULT 1  
AF039699 2162 bp mRNA PRI 05-MAY-1999  
LOCUS Homo sapiens antigen NY-CO-37 (NY-CO-38) mRNA, complete cds.  
DEFINITION AF039699  
ACCESSION AF039699.1 GI:3170197  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2162)  
Scallan,M.J., Chen,Y.T., Williamson,B., Gure,A.O., Stockert,E.,  
Gordan,J.D., Tureci,O., Sahin,U., Pfeundschuh,M. and Old,L.J.  
Characterization of human colon cancer antigens recognized by  
autologous antibodies  
JOURNAL Int J Cancer 76 (5), 652-658 (1998)  
MEDLINE 98272252  
REFERENCE 2 (bases 1 to 2162)  
Scallan,M.J.  
AUTHORS Direct Submission  
TITLE Submitted (23-DEC-1997) Ludwig Institute for Cancer Research, New  
JOURNAL York Branch at Memorial Sloan-Kettering Cancer Center, 1275 York  
Avenue, New York, NY 10021, USA  
FEATURES  
source  
1. 2162  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="colon cancer metastasis to lung"  
1. 2162  
/gene="NY-CO-38"  
97. .1308  
/note="pdz domain protein; NY-CO-38 antigen splice  
variant"  
/codon\_start=1  
/product="antigen NY-CO-37"  
/protein\_id="AAC18048.1"  
/db\_xref="GI:3170198"  
/translation="MDRKVAREFRHKVDPLIENDAKDYLDVLMYHOTMDVAVLVG  
DKLYINERSRLPLDARPLPLKHQVYDQYDETVRLNGSISCTHEEYINLRKTSV  
RGLEFGGLFISHLIKGGQADSVGLQYDETVRLNGSISCTHEEYINLRKTSV  
SLKRVHGLIPVKSXPDEPLTMQYDQVSESGVSGLSGSGNENNEKVFISLVG  
SRGLCSISSGPIQPGIETSHVKPGSLSAEVGLIGDIYEIVNGVDNSLDHKEAVN  
VKNRSRLSIVAAAGRELFWTDRELRSLAEVGLIGDIYEIVNGVDNSLDHKEAVN  
OEMERORREIKAKAEENERYRKEMQIVVEEEKFKQMEWDSKQDLLPKTITA  
EYHVPYLRKPK"

Query Match 100.0%; Score 2162; DB 40; Length 2162;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 cctggccggtcgcggtcgcggtcttcacagtcctcggcagcgccgagcccgaaagaac 60  
DB 1 CCGGGCCCGGTGCGCGGTCTTCCAGCTCTGCGAGCCGGGACCCGGAAGAAC 60  
QY 61 gggctgtcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 120  
DB 61 GGGTCTGCAACGACGACCTGACCTGCGCCAGCCATGAGCCGAAAGTGGCCAGAA 120  
QY 121 ttccgcataaagtgatattcttgatgtaaaatgatacgcagagagagactatctatgat 180  
DB 121 TTCGGGCAATGAGTGATTTTCTGATTGAATGATGCAAGAGACTATCTCTATGAT 180  
QY 181 gtcgtgcaatgtaccacacagacatgacgtgacgtgtcgtgtggaagacctgaaactg 240  
DB 181 GGTGCTGCAATGTACCAACAGACGATGACGTGGCCGTCTGTGGAAGACCTGGAAGCT 240  
QY 241 gtaacataaagcagcagcgtctgctctgttggatgcatcgcgcgtgatccactg 300  
DB 241 GTCATCAATGAACCCAGCCGCTGCTGTTGATGCCATTGCGCGCTGATCCACATG 300  
QY 301 aagcagcagtgatgataatgatcagctgaccccccgccgctccaggaagctgaaagagtg 360  
DB 301 AAGCACCAGGTGGAATATGATATGCTGACGACCCCGGCGCTCCAGAGAGCTGAAGAGGTG 360  
QY 361 cgtctggaacgtctgcaccccggaagcctcgccctgagtgagtgagtgagtgagtgag 420  
DB 361 CGTCTGACCGTCTGACCCCGGAGGCGCTGAGTGCGTGGTGGCTGGCGAGTGT 420  
QY 421 ggtctgaggtcttcatctccacccatcaaaaggggttcaggaagcagcagcagcagcag 480  
DB 421 GGTGTTGGGCTCTCTATCTCCACCTCATCAAGGGGCTGAGGCAACAGCGTGGAGCTC 480  
QY 481 caggtgaggaagagatctgcagatcaatgataatccatccctccctgataccatgag 540  
DB 481 CAGGTAGGGGACAGATCTGTCGGATCAATGATATTCATCTCTCTGATCCATGAG 540  
QY 541 gaggatcaacatcattcgaaacaaagaatgltgcatcaaaagtgaacacatcagcagc 600  
DB 541 GAGGTCAATCAACCTCATTCGAACCAAGAAAGCTGTCCATCAAAAGTAGACACATCGCG 600  
QY 601 ctgattcccggtgaaagcctctctgataagccctacttgagcagtgatgtgagtgatt 660  
DB 601 CTGATCCCGGTGAAGAGCTCTCTGATGAGCCCTCACTTGGAGATGTGATGATGATG 660  
QY 661 gttggaatctggggggtgaggaagcagcctgaggtccctcggaatctgggaaacaaag 720  
DB 661 GTGTGCGAATCTGGGGGCTGCGAGGACACCTGGCTCCCTTGGAATGGGAAACAAAG 720  
QY 721 gagaagaaggtcttcacagcctgtagcctcccgagagccttgctgtagcagcatlccagc 780  
DB 721 GAGAAAGAGTCTTTCATCAGCCTGTGAGGCTCCGAGGCTTGCGTGCAGCATTTCTCAC 780  
QY 781 ggcgccatccagaagccttgatcttatacagcactgtgaacactggtccctgtctgct 840  
DB 781 GGGCCCATCAGAAAGCTGTGCATCTTATCAGCCAGTGAACCTGGCTGCTGTCTGCT 840  
QY 841 gagggtgagttggaataaggaagcagatgtgtcgaagtaagtgaggtcgtactcttaac 900  
DB 841 GAGGTGGGATTGGAGATTAAGGAGACAGATGTTGCAAGTAAATGCGCTCACATCTCTAAC 900  
QY 901 ctggaatcaaaaggaagctgtgtaaatgtctgtaaaaaatagccgaagcctgacatctccat 960  
DB 901 CTGGATCACAAGAGGCTGTAAATGTGCTGAATAAATAGCCGACGCTGACCATCTCCATT 960  
QY 961 gtaagctgcaagcttgccgggaagctgttcatgacagacgcgggaagcgtgcaagcgcg 1020  
DB 961 GTAGCTGACAGCTGGCGGAGAGCTGTTCATGACAGACCGGAGACGCGTGCAGAGCGCG 1020  
QY 1021 cagcgtgagctcagcggcaggaagccttcacatgcagaagcgtgagcgtgagtgagtcac 1080  
DB 1021 CAGCGTGAAGCTGCGAGGCGGAGAGCTTCTCATGACAGACCGGCTGCGCATGAGTCCAAC 1080  
QY 1081 aagatcctccaggaagcagcaggaagatgtagcggcgaaggaagaagaatgtgccagaag 1140  
DB 1081 AAGATCTCTCAGAGAGCAGCAGAGATGAGCGCAAAAGAGAAATGTCGCCGAGAG 1140  
QY 1141 gcaagcaggaagaatgagagataccggaaagagatggaacaatattgtaaggaaggaag 1200  
DB 1141 GCAGCAGAGAAATGAGAGATACCGGAAGAGATGGAACAATGTAAGAGAGGAAGAG 1200  
QY 1201 aagtttaagaagaacaatggaagaagactggggtcacaaggaacagctactctgtctaa 1260  
DB 1201 AAGTTTAAGAACAATGGAAGAAGACTGGGGCTCAAAGAACACACTACTCTTCCCTTAA 1260  
QY 1261 accatcactgtcagagttacaccacagtaccccttcgaagccaagtgtattccggaaba 1320  
DB 1261 ACCATCACTGTGAGAGTTACACCAGTACCCCTTCCGAACCAAGTGATTTCCGGAATA 1320  
QY 1321 tgaaggaaggtcttgacccctactatgttcaaccccaagagacagatcaggggaagatgt 1380  
DB 1321 TGAAGGAAGGTCTTGACCCCTACTATGTTCACCCCAAGAGACAGATCAGGGGAAGATGT 1380



Db 421 GGCCTGGGCTCTTCATCTCCACCTTCATCAAAAGCGGCTGAGGACAGACGCTGGGCTC 480  
 QY 481 cagctagagagacagatcgtccgatacaatgatatccatctccctctgtaacatgag 540  
 Db 481 CAGCTAGAGGACAGATCGCCGGATCAATGATATTCCATCTCTCTCTTACCCTATGAG 540  
 QY 541 gaagtcatacaactcattcgaaacaaagaaactgtgtcacaatgaaatgagacatcgtgc 600  
 Db 541 GAGCTCATCAACCTCATTCGAACCAAGAAACGTGTCTCATCAAGATGAGACATCTGGC 600  
 QY 601 ctgatacccgtagaaagctctctgtatgagccccctcaactgtgagatgtgtgagatgtt 660  
 Db 601 CTGATCTCCCTGAAAGACTCTCTGTGATGAGCCCTTCATCGGACGATGTGATGATGATTT 660  
 QY 661 gtctcgaaatctggggcgctgagagcaagcttgagctccctcgaaatcgagaaacaag 720  
 Db 661 GTCTCGGAATCTGGGGCGCTGGCAGGACGCTGGGCTCCCTTGGAAATCGGGAACAAAG 720  
 QY 721 gagaagaaggtctcattcagcctgtgtaggtcccgagagccttggtctgagcatctccagc 780  
 Db 721 GAGAAAGAGGCTTCATCAACCTGTGATGAGCTCCGAGGCTTGGCTGACACATTTCCAGC 780  
 QY 781 ggcacctcacaagaagctgcatcttatacagcatgtgaaactgtgtccctgtctgtc 840  
 Db 781 GGCCTCATCCAGAGCTGAGCATCTTATCATGACCATGTGAACCTGCTCCCTGCTGCT 840  
 QY 841 gaagctgagattgtagatagagagcaagatgtctgaatgcaatgagctgcagctctctaac 900  
 Db 841 GAGCTGGAGTGTGAGATAGGGGACAGATTTGTGAATGCAATGGCTGACATCTCTTAAC 900  
 QY 901 ctgatacaacaaggagctgttaaatgtgtcgtgaaataagccgacagctgacatctccatt 960  
 Db 901 CTGATCAACAAGAGGCTGTAAATGTCTGAAAAATAGCCGACCTGACATCTCATTT 960  
 QY 961 gtatgcctgagctgtgcccggagagctgttcatagacagaccggagagcgctgtgagagcgcg 1020  
 Db 961 GTATGCTGACAGCTGGCCGGAGCTGTTCATGACACCGGAGGCGGTGGAGAGGGCGG 1020  
 QY 1021 caagctgagctgtgagcggagagctctcattcagaagcggtgtgagatgagctaac 1080  
 Db 1021 CAGCTGAGCTGTGAGCGGACGAGACTTCTCATGACAGAGCGGCTGGCGATGAGTCCAAC 1080  
 QY 1081 aagatctctcagaagagcagcagagatgtgagcgcaagagaaagaaatltgcccagaag 1140  
 Db 1081 AAGATCTCTCAGAGAGCAGCAGAGATGAGCGGCAAGAGAAATTTGCCAGAAG 1140  
 QY 1141 gaagcagagaaatagagatataccgaaagagatgtgaaacagattgtagaagagagag 1200  
 Db 1141 GCAGCAGAGAAATAGAGATACCGGAAGAGATGAAACAGATTGTAGAGAGAGAGAG 1200  
 QY 1201 aagtttaagaagaacatgaggaagaagactggggtcacaagaacagctactctgtctaa 1260  
 Db 1201 AAGTTTAAAGACAATGGGAGAGAACTGGGCTCAAGAAACAGCTACTCTTGCTTAA 1260  
 QY 1261 accatcaactgtgtgaggtacacacagtaaccctcgtgaagccaaagt----- 1307  
 Db 1261 ACCATCACTGTGAGGTACACCCAGTACCCCTTCGCAAGCAAAATGTATCATGAGAGTGT 1320  
 QY 1307 ----- 1307  
 Db 1321 GAACCTGAGCTCGAGCCCGCAGATGACCTGTGATGAGGACGAGAGAGAGAGAGAGAG 1380  
 QY 1307 gattccggaataataggaagaagcttgagaccctactctatgttcaacccagaagagatc 1366  
 Db 1381 GATTTCGGGAATATAGAGAGAGGCTTTGACCCCTACTCTATGTTCACCCAGAGCAGATTC 1440  
 QY 1367 atgtgggaagagatgtccggtcctcagcatcaagaagaagagagatccttagaactgtgcctg 1426  
 Db 1441 ATGGGGAAGAGATGTCCGCTCTTACCATCAAGAAAGAGAGATCTTAAAGCTGCGCTTG 1500  
 QY 1427 gaagcggtgtgtgagctcccccattgtggaaggtgtgtctgtctgtctgtatgagcgagga 1486  
 Db 1501 GAAGCGGCTGTGAGACATCCCATTTGGGAAGGTGTGTGTCTGTGTATGAGCGGGA 1560

QY 1487 gctctgagcggcatgtgtgcatctgtgaaagggagacagatcagtgcaatcaacggcaag 1546  
 Db 1561 GCTGCTGAGGCGGATGTGGCATTTGTGAAGGGGAGAGATCTATGGCAATCAAGGGCAAG 1620  
 QY 1547 attgtacagatacaccctgtgtgagctgtgagctgtccctgcagaagggcttgaatcag 1606  
 Db 1621 ATTGTACAGATACACCCCTGTGAGGCTGTGAGCTGTGCCCTGCAGAGGCTTGAATTCAG 1680  
 QY 1607 ggcgggagactgagatgagctgtgtgtgtccgtctgcccccaagagatgagcgtgag 1666  
 Db 1681 GGCGGGACTGTGATGACCTTGTGTGGCTGTGCCCTCCCAAGAGATGACGATGAGAG 1740  
 QY 1667 ctgacctctgtctgtagatccaaagggagaaacaaatcaacgcgttaagaaacaatgag 1726  
 Db 1741 CTGACCTCTGTGCTGAGATCCAAAGGGAACCAATTCACCGCTTAGAAACATGTGAG 1800  
 QY 1727 ctccggcccaactcgttgaacacaaagcctcggacagccttgagagagggccacatgaca 1786  
 Db 1801 CTCCGGCCCACTCGTGAACACAAAGCCCGGACAGCCTTGAGAGAGGCCACATGACA 1860  
 QY 1787 cacaccagatgagatccttgaggacctgaaatctatacccaagagatctcaactcccttg 1846  
 Db 1861 CACACGAGATGSCATCTTGGGACTGTGAATCTATCAACCAAGAGATCAAACTCCCTTG 1920  
 QY 1847 gcccgaacaaggccagataaagaaacagctcgagccacttttgaagggcaatgtgga 1906  
 Db 1921 GCCCTAACCAGGCGCAAGATTAAGAAACAGCTCGGGCCTTTTGAAGCCCAATGTGCA 1980  
 QY 1907 ggaagggagcagcagcggttggagagaagatcgaagaatcagaactcactcctt 1966  
 Db 1981 GGAAGGGAGCAGCGCCGTTGGGAGAAATCTCAAGATCAACACATCTCATCTCTTT 2040  
 QY 1967 cctctggccaagtgaattgtgtctcctcagactgtggggagctcctcttgaacctta 2026  
 Db 2041 CCTTGCCCAAGTAATTTGTCTCTCCAGCTTTGGGAGCTCTTCTTGAACCTTAA 2100  
 QY 2027 taagaccacactgagatctctctctcattccatccctcctcctgcctcgtctaaatgc 2086  
 Db 2101 TAAGACCCACATGAGATCTCTCTCTCATCCCTCTCTCTGCTGCTGCTTAATTCG 2160  
 QY 2087 tgcagagattgtcactcacaaccttactctgtgagctatcaataataaataacagatttatt 2146  
 Db 2161 TGCCAGAGATTTGATCACTCAAAACCTTACTGTGAGCTCATTAATTAATAACGATTATT 2220  
 QY 2147 ttccagcttaaaaaa 2162  
 Db 2221 TTCAGCTTAAAAAA 2236

RESULT 3  
 AB018687  
 LOCUS  
 DEFINITION Homo sapiens mRNA for autoimmune enteropathy-related antigen AIE-75, complete cds.  
 ACCESSION AB018687  
 VERSION AB018687.1 GI:5231270  
 KEYWORDS autoimmune enteropathy-related antigen AIE-75; AIE-75.  
 SOURCE Homo sapiens small intestine cDNA to mRNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 REFERENCE  
 AUTHORS  
 TITLE  
 Kobayashi, I., Immura, K., Kubota, M., Ishikawa, S., Yamada, M., Tonoki, H., Okano, M., Storch, W. F., Moriuchi, T., Sakiyama, Y., and Kobayashi, I.  
 2 (bases 1 to 2213)  
 Kobayashi, I.  
 Direct Submission



OY	1614	actgagatgcaccttgggttgccgctcgcccccaaggagtatgaagatgatgcacct	1673
Db	1681	ACTGGAATCGACCTTGTGGTTGCCGTGCCTCCGCCCAAGAGATATGACGATGACTGACCT	1740
OY	1674	tcttgttaagttccaagaagggaaccacaatttcacggtttaaggaaacagttagctccggc	1733
Db	1741	TCTT-CTGAAGTGCCAAAGGGGAAACCATAATTTCAC-GCTTAGGAAACAGTAGCTCCGGC	1738
OY	1734	cccacctcgltyaacacaaaagcctcgcagccagcctltgagagaggccaatgacacacacca	1793
Db	1799	CCCACCTCGTAACACAAAAGCTTCGGATFCAGCCTTGAGAGAGGCGCACACTACACACACA	1858
OY	1794	gatggcatccttgygagactgaaatctataccaccaggaattccaactccttggccctga	1853
Db	1859	GATGCATCTCTTGGAGCATGAACCTATACACCCAGGAATFCAAACTCCCTTTGGCCCTGA	1918
OY	1854	accagggccagataaagaagacgtcggcgcaacttttttgaaggccaatgtgaggaaagg	1913
Db	1919	ACCAGGGCCAGATPAAGAACAGCTCGGGCCACACTTCTTGAAAGGCAACGTGAGAAAGG	1978
OY	1914	gagcagccagccglttggggaagaatctcaagaatccagaactcattccttctctctg	1973
Db	1979	GAGCAGCCAGCCATTGGGAGAAGTCTCAAGATCCAGACTCTCATCTCTTCCCTCTCG	2038
OY	1974	cccagtgaaattggtctctcccaagcttggggagactccttccttgaacctataagacc	2033
Db	2039	CCCAGGAATTTGGTCTCTCCAGCTCTGGGGGAGCTCTTCCTGAACCCCTAATPAAAGCC	2098
OY	2034	gcactggaatctctctctccalcactctccctcctgcctcctctctaattgctgcagag	2093
Db	2099	CCACTGGAATCTCTCTCTCCATCCCTCTCTCTGCCCCCTCTTAATTGCTGCCAGG	2158
OY	2094	attgtcacctccaaccttact	2114
Db	2159	ATTGTCATCTCCAACCTTACT	2179
RESULT	4		
AB006955	LOCUS	AB006955 2283 bp mRNA	PRI 02-DEC-1999
DEFINITION	Homo sapiens mRNA for AIE-75, complete cds.		
ACCESSION	AB006955		
VERSION	AB006955.1 GI:5152287		
KEYWORDS	AIE-75; aie-75.		
SOURCE	Homo sapiens small intestine cDNA to mRNA.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homidae; Homo.		
	1 (sites)		
	Kobayashi,I., Imamura,K., Kubota,M., Ishikawa,S., Yamada,M.,		
	Tomoki,H., Okano,M., Storch,W.F., Moriuchi,T., Sakiyama,Y. and		
	Kobayashi,K.		
TITLE	Identification of an autoimmune enteropathy-related 75-kilodalton		
JOURNAL	antigen		
MEDLINE	Gastroenterology 117 (4), 823-830 (1999)		
REFERENCE	99431904		
AUTHORS	2 (bases 1 to 2283)		
TITLE	Kobayashi,I.		
JOURNAL	Direct Submission		
	Submitted (31-AUG-1997) to the DDBJ/EMBL/Genbank databases. Ichiro		
	Kobayashi, Hokkaido University School of Medicine, Department of		
	Pediatrics, North-15, West-7, Kita-ku, Sapporo, Hokkaido 060-8638,		
	Japan (E-mail:ichikoba@med.hokudai.ac.jp, Tel:011-716-1161,		
	Fax:011-736-9267)		
FEATURES			
Source	Location/Qualifiers		
	1..2283		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/tissue_type="small intestine"		
gene	246..1811		
	/gene="aie-75"		
CDS	246..1811		

Query Match	85.3%	Score 1843.6	DB 9	Length 2283
Best Local Similarity	94.7%	Pred. No. 0		
Matches 1994	Conservative	0	Mismatches 29	Indels 83; Gaps 5
BASE COUNT	549 a	594 c	670 g	470 t
ORIGIN				
Query Match	/gene="aje-75"			
	/codon_start=1			
	/product="Aie-75"			
	/protein_id="BAA81739.1"			
	/db_xref="GI:5152288"			
	/translation="MYHOMDVAVLVGDLKLVINEPSRLPLEDAIRPLIKHOVEYET OUTPRRSRIKIEVRLDRDLRAPEIGLSVSGGELFGLGLISLIRKGGADSVGLGDP IVRINGYISLSTTHEEVLINLIRTKTVYIKRHHIELIPKSSPDEPLTWQYVDYFVS SGGVGLSGPGRNRENKRVKIVLSVGRSGLSSTSP IOKPGLFISHVPGSLSA VGLIGDQIIVEYVDFPSNLDHKEAVNYLVNRSRLTILIVAAQELHMTDRLETE RORELRQBELLMOKRLAMESNKILQEOEMERORAKETIAOKAAEENERYRKMEQIYV BEERKKOMEEDMGSKBOLLKPTTAAVHPPLPKPYDQGVPELEPPADLGGTE BOGEDPFRKYEBGEPDPYSMFTPEQIMGDVRLIRKKGSIDLAEQVDSPIKVVYV SAPVIRGAERHGGIVLKQDEIMAILNKIVTDTTIAEAPALQKAMNQGDDILVAVAA CPKREYDDELTFE"			
131 aggtgattttctgattgtaaatgatagcagaagaagacattctctatgtgtgcgcgaa	190			
187 AGGTGATTTTCTGATTGTAATAATGATGACGAGAAAGACATCTCTAATGATGCTCGCA	246			
191 tgcaccacagacacatgacgcgtgcgcgtctcgttggagaacctgacatgcatcaatg	250			
247 TGTACACACGACACATGACGACGTCGCCGTCTCTGTGGAGACCTGAAGCTGCTCATCAATG	306			
251 aaccacacccgtctgcctctgtttgatgacatcgcgcgtgatalcccaatgaagcaaccag	310			
307 AACCCAGCGGTGTGCTCTGCTTGTGATGACCATTCGGCCGCTGATCCACATGAAGCACACAG	366			
311 tggaaatgatatacgtgacccccccgcgtccaagaagtgtgaagagtgctcggacc	370			
367 TGGAAATATATATAGCTGACACCCCCCGCGCTCCAGGAAGCTGAAGAGTGCTCTCGAC	426			
371 gctcgaccccgaaagccctcgagctgagctgctggtgagcttggaatttgctctgagac	430			
427 GCTGTCACCCCGAAGGCCCTCGGCTGAGTGTAGTGTGAGTGTGCTGGAAGTTTGGCTGTGGCC	486			
431 tcttcatactccacactcatcaaaagcgtgcaagcagacacgctcgggctccagtagaggg	490			
487 TCTTCATCTCCACCTCATCAAAAGGGGGTCAAGGACAGACGCTCGGGCTCCAGTAGAGGG	546			
491 acgagatgctccgatatcaatgtatatccattctctcctctgtaccatgaagagatcatca	550			
547 ACGAGATCTCTCGGATCATATGATATTCATCTCTCTCTGATCCCATGAGGAGGTATCATCA	606			
551 acctcaatcgaaacaagaanaactgtgtccataaaagtgtgaagacaactcgagccgatccccg	610			
607 ACCTCATTCGAAACCAAGAAACTGTGTCCATCAAAAGTGAACATCGGCTGTATCCCGC	666			
611 tgaaaagctctctctgatgagccctcccaactctggcagtatgtgatacagtttgttggaaat	670			
667 TGAAGAGCTCTCTCTATATAGCCCCCTCACATTTGGCAGATGTGATACAGTTTGTGTGGAAAT	726			
671 ctgggggctgtcgaagcagccttggtctccctctgaaatcgggaaaaaagaagaaaggg	730			
727 CTGGGGGCTGTGGAAGCAGCCCTTGGGCTCCCTCGGAATCTGGGAAAAAAGAGGAAGAGG	786			
731 tcttatacagcctgttagctctccgaagccttggtgtgaagcattccagcgagccccatcc	790			
787 TCTTATATCACTGGTGTAGGCTTCCAGAGGCTTGGCTGACATTTTCCAGGGCCCCCATCC	846			
791 aagaagcctgcgaactcttatacagcactgtgaaaccctggctccctctgtctgaagtgagat	850			
847 AGAAGCTTGCGATCTTTATCAGCCATGTGAAACCTTGGCTCCCTGTCTGTGAGTGGGAT	906			
851 tggagataggggacagattgtcgaagtaacatgagctgcgacttcttaacctgtacaica	910			

D	b	907	TGAGATATAGGGGACCAACATATGTCGAAGTCAATGGCTCCACATTCTCTATACCTGGATGCACA	966
O	y	911	aggagcctgttaaatlytgcctgaaaaataatagccgcagcccttgaccatctccatctgtatgtcag	970
D	b	967	AGGAGCGCTGTAAATGTCTGTAAGAGTATGCCGACGCTGCACATCTCCATTTGATGTAGCTGCAG	10266
O	y	971	ctgagccgggagctgttcaatacagaccgggagcgcgtctgycagaagcgcgagcgtagc	10300
D	b	1027	CTGGCCGGGACCTGTTTCATATACAGACCCGGAGGGCGGTGACAAGCGCGGCGGCGCTGTAGC	10866
O	y	1031	tgcaagccgaagagcttctcatatgcagaagggctggtgcgaatggatctcaaaacaagaatccctcc	10900
D	b	1087	TGCACGGCGACGAGACTCTTCATATCAGAGCGGCTGGCGATGAGTCCACACAGATCTCTCC	11466
O	y	1091	aggagcagcagagatgtgagccgycgaagaagaaaaaattgcccagaagcgacagag	11500
D	b	1147	AGGAGCAGACAGAGATGTAGCGGGCAAAAGGAAATAATTCGCCAGAAAGCAGCAGAG	12060
O	y	1151	aaatatgagagataccggaagagagatgtgaaacagatctgtagaagagagaagaagtttaaga	12100
D	b	1207	AAATATGAGATATACCGGAAAGAGATGTGAACAGATTTGTATAGAGAGAAAGAAATTTAAGA	12660
O	y	1211	agcaatgggaagagagacttggggtccaaagaagaagctactcttgctcttaaaacactacgt	12700
D	b	1267	AGCAATGGGAAGAAAGACTGGGGCTCAAAAGAGACAGCTACTTTGCCCTAATAACATCATCTG	13260
O	y	1271	ctgaggtacaccccaqtaaccctctgcgaagccaaagt-----gattccgga	13070
D	b	1327	CTGAGGTATACACCGGTACCCCTTCGCAAGCCAAATATCATCAGGGATGTGAACCTGAGC	13860
O	y	1307	-----gattccgga	13160
D	b	1387	TGCGAGCCCGAGATGACCTGGATGTGAGGCGACGGAGGAGAGGAGAGCAGATTTCTCCGA	14460
O	y	1317	aaatggggaagaggtttgaccctctctatgttcaaccacaagaagatctatggggaag	13760
D	b	1447	AATATGAGAAAGGCTTTGACCCCTACTCTATGTGTCAACCCCAAGCAGATCATGGGGAAAG	15060
O	y	1377	atgtccgcgtctccatcgcatcaagaagagagatcccttaagcttgcgccttgagaagccggt	14360
D	b	1507	ATGTCCGGCTCTTACGCATCAAGAAGAGGAGATCTTAAACCTGGGCGCTGGAAAGCGGCTG	15660
O	y	1437	tggaactccccaattgggaaggtgtgcgttctgtctgtatgaagcgggagcctgcctgagc	14960
D	b	1567	TGGACTCCCCCATTTGGGAAGGTGTGTCTTCTGCTGTATGAGCGGGAGCCTGCTGAGC	16260
O	y	1497	ggcctgtgtgcatgttgaagaaggagcagagatcatgtgcaatcaagcgcaagatgttgyaag	15560
D	b	1627	GGCATGGGTGATTTGTAAAGGGGAGCAGATCATGGCAATCAACGGCAAGATTTGTAGAG	16860
O	y	1557	actaacaccctgtgcagagcttgcagccttgccttgagaagggccttgcagggccggagct	16160
D	b	1687	ACTTACACCTTGCTGAGGCTGTAGGCTGTCCCTCGAGAAAGCGCTTGGAATCAGGGCGGGACT	17460
O	y	1617	ggatcgaacctgtgttgcgcgtctgcgcctcccaagagatgacgtatgcgttgcacctct	16760
D	b	1747	GGATCGACCTTGTGTGGTGGCGTCTGCCGCCCAAAAGAGATATGACGTATGAGCTGACTTCT	18060
O	y	1677	tgctcgaagtcacaagaagggaaccaaatatcagcgtttagaagaacgttgcgtccggcccc	17360
D	b	1807	T-CTGAAGTCCAAAAGGGGAAACCAATTCAC-CTGTAGGAACAGTGTAGCTCCGGCCCC	18460
O	y	1737	accctgtgaacacaagaagccctgcgaaccagccttgaagaagggcacatgaacacaacagat	17660
D	b	1885	ACCTCTGTGAACACAAACCTCTCGGATCAGCCTTATAGAGAGGCCACACTTACACACCAAT	19240
O	y	1797	ggcatccttgggacctgaatctatcaaccaagaatctcaaatcccttcttgccttgaac	18560
D	b	1935	GGCATCTTGGGACCTCAATCTATTCACCCAGGAATCTCAACATCCCTTTGGCCCTGAAC	19840
O	y	1857	aggcgccagataagaacaagcttgggcacacttttttgaagggcaatgltgtgaaagaaggag	19160
D	b	1985	AGGGCCAGATTAAGGAACAGCTCGGGCCACTCTTCTAAAGGCCAACGCTGGAGGAAGGGAG	20440

QY	1917	cagccagccgctttgggaggaagctctaagaatccaaagactctatcttcctctcgagcc	1976
Db	2045	cagccagccatttgggaggaagatctcaagatccaaagctctatctcttctcgagcc	2104
QY	1977	agtaatttgctctctccagactttggggagctcctctctttaaaccataaagacc	2036
Db	2105	agtaatttgctctctccagactttggggagctcctctctttaaaccataaagacc	2153
QY	2037	ctggaagctctctctccactcctctctctctgagcctgcctctaatgctggcagagt	2096
Db	2164	ctggaagctctctctccactcctctctctctgagcctgcctctaatgctggcagagt	2233
QY	2097	gtcaactcnaaacttactctgagctcatatataaataaacagatatttccagcta	2156
Db	2234	gtcaactcnaaacttactctgagctcatatataaataaacagatatttccagcta	2277
QY	2157	aaaaa	2162
Db	2278	aaaaa	2283

RESULT		5		PRI		30-SEP-1988
AC005137/c						
LOCUS	AC005137	161190 bp.	DNA			
DEFINITION	Human Chromosome 11p14.3 PAC clone 6-106f23, complete sequence					
ACCSSION	AC005137					
VERSION	AC005137.1	GI:3687273				
KEYWORDS	HTE,					
SOURCE	human,					
ORGANISM	Homo sapiens					

REFERENCE  
1 (bases 1 to 161190)  
AUTHORS  
Evans, G.A., Athanasiou, M., Aguayo, P., Armstrong, D., Basil, M.,

**TITLE**  
HHS Submission

unpublished  
2 (bases 1 to 161190)  
Evans, G.A., Athanasiou, M., Aguayo, P., Armstrong, D., Basil, M.,

**TITLE** Direct Submission  
**JOURNAL** Submitted (19-JUN-1998) Genome Science & Technology Center,  
Buetner, J., Bumeister, R., Card, P., desallboal, C., Dunn, J.,  
English, C., Ehrhardt, S., Garner, H. R., Gee, V., Gordon, M., Gotway, G.,  
Grant, O., Hahner, L., Harris, J., Lewis, E., Loo, H., Loo, K. N.,  
Major, T., McFarland, J., Newton, J., Osborne-lawrence, S.,  
Schageman, J., Schultz, R. A., Stimson, S., Syed, M. and Ward, T.

REFERENCE  
3 (bases 1 to 161190)  
University of Texas Southwestern Medical Center, 5323 Harry Hines  
Blvd, Dallas, TX 75235-8591, USA

**AUTHORS**  
Evans, G.A., Athanasiou, M., Aguayo, P., Armstrong, D., Basit, M.,  
Buettner, J., Butler, C., Card, P., desilboet, F., Dunn, J.,  
English, C., Ethridge, S., Garner, H.R., Gee, V., Gordon, M., Gotway, G.,

Grant, O., Bahner, L., Joslin, J., Lewis, E., Ioo, H., Loo, K.N., Major, T., McFarland, J., Newton, J., Osborne-Lawrence, S., Schageman, J., Schultz, R.A., Stimson, S., Waller, K. and Ward, T.  
 Direct Submission  
 Submitted (30-SEP-1998) Genome Science & Technology Center,  
 University of Texas Southwestern Medical Center, 5323 Harry Hines  
 Blvd, Dallas, TX 75235-8591, USA  
 On Oct 1, 1998 this sequence version replaced gi:3241950.

Further information regarding the map of this region or annotation of 6-106f23 can be found at <http://gdb.stec.su.med.edu/chromosome2.htm>. CHROMOSOMAL LOCUS: This PAC clone maps from the Usher syndrome region (USH1C) mapped between STS markers D15S1310 and 115S14. MARKER CONFIRMATION: STS/ESTs sequence confirmed; D15S1032 and D15S1318. MAPPED CLONE OVERLAP: pDJ239b22 and 6-109h6.

IMPORTANT: This submission contains the entire insert of clone 6-106f23. 6-106f23 comes from a PAC library constructed at the Roswell Park Cancer Institute by the Pieter de Jong group. This clone has been finished according to strict quality criteria and attempts have been made to resolve all base calling problems such as compressions and repetitive elements. The expected Phred/Phrap calculated errors/10kb is 0.40. In addition, this sequence has been finished such that 99.9% of consensus base calls consist of either double-stranded coverage or 2 types of labeling chemistry on one strand.

## FEATURES

```
Source      Location/Qualifiers
repeat_region 1..161190
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              574..955
              /rpt_family="MT2B2"
repeat_region 1445..1635
              /rpt_family="MIR"
              3125..3302
              /rpt_family="MIR"
              11175..11329
              /rpt_family="MIR"
              14552..14716
              /rpt_family="MIR"
              complement(18589..18627)
              /rpt_family="MIR"
              23165..23232
              /rpt_family="MIR"
              complement(25103..25361)
              /rpt_family="Alu"
              29622..29909
              /rpt_family="Alu"
              complement(32772..32989)
              /rpt_family="Alu"
              33175..33264
              /rpt_family="MIR"
              complement(34762..34852)
              /rpt_family="MIR"
              34864..34996
              /rpt_family="MIR"
              complement(35719..35843)
              /rpt_family="Alu"
              complement(36294..36467)
              /rpt_family="MIR"
              36817..37084
              /rpt_family="Alu"
              37953..38186
              /rpt_family="MSR1"
              39000..39177
              /rpt_family="MIR"
              complement(39898..39968)
              /rpt_family="MIR"
              complement(41579..41780)
              /rpt_family="MIR"
              complement(43317..43597)
              /rpt_family="Alu"
              45609..45899
              /rpt_family="Alu"
              complement(46819..46884)
              /rpt_family="MIR"
              complement(48840..48895)
              /rpt_family="MIR"
              50385..50543
              /rpt_family="MIR"
              complement(50842..51126)
              /rpt_family="Alu"
              complement(53140..53180)
              /rpt_family="MIR"
              complement(55758..56060)
              /rpt_family="Alu"
              complement(59724..60021)
              /rpt_family="Alu"
              66550..66645
```

```
repeat_region 73180..73260
              /rpt_family="MIR"
              complement(74694..74990)
              /rpt_family="Alu"
              complement(76214..77037)
              /rpt_family="MIR"
              complement(77130..77414)
              /rpt_family="MIR"
              complement(77532..77641)
              /rpt_family="MIR"
              complement(78518..78630)
              /rpt_family="Alu"
              complement(78918..79211)
              /rpt_family="Alu"
              85547..85650
              /rpt_family="MIR"
              88499..88613
              /rpt_family="MIR"
              complement(93509..93900)
              /rpt_family="Alu"
              complement(94148..94417)
              /rpt_family="Alu"
              complement(95536..95822)
              /rpt_family="Alu"
              complement(96013..96567)
              /rpt_family="Alu"
              complement(106391..106583)
              /rpt_family="MIR"
              complement(114951..115136)
              /rpt_family="MIR"
              118500..118580
              /rpt_family="MIR"
              complement(126403..126593)
              /rpt_family="Alu"
              126665..126942
              /rpt_family="Alu"
              complement(127152..127352)
              /rpt_family="Alu"
              complement(128074..128639)
              /rpt_family="Alu"
              complement(128923..129196)
              /rpt_family="Alu"
              130009..130278
              /rpt_family="Alu"
              complement(134144..139342)
              /rpt_family="Alu"
              complement(137205..137469)
              /rpt_family="Alu"
              complement(140667..140762)
              /rpt_family="MIR"
              complement(142407..142521)
              /rpt_family="MIR"
              147830..148006
              /rpt_family="MIR"
              149821..150066
              /rpt_family="Alu"
              complement(151137..151327)
              /rpt_family="MIR"
              160309..160771
              /rpt_family="MIR"
              BASE COUNT 37674 a 40113 c 40195 g 43208 t
              ORIGIN
```

Query Match 20.2% Score 437.4; DB 11; Length 161190;  
Best Local Similarity 95.7%; Pred. No. 9.9e-91;  
Matches 472; Conservative 0; Mismatches 16; Indels 5; Gaps 2;

QY 1666 gctgacctcttgcctgaagtcacaaagggaacacaaattcacgcgttagagaaacagtcga 1725  
DB 5253 GGAGACCTCTTCTGCAAGTCCAAAAGGGGAAACCAATTCAC -CGTTAGGAACAGTGA 5195



Oy	1726	gctccggccccaactctcgtgaacaacaagccttcggaaccaagctctgaaggagggccacatgac	1785
Db	5194	gctccggccccaactctcgtgaacaacaagccttcggaaccaagctctgaaggagggccacacactac	5135
Oy	1786	acaacacgaatggaactccttggaacctgaactctaacacacgaagaatccaactcccttc	1845
Db	5134	ACAACAACGATGGCATCTCTTGGGACCTGAAATCTATCAACCACGGAATCTCAAACTCCCTTT	5075
Oy	1846	ggccctgaaccagagggccagataaagaaagcctcgggccactttttgaagggccaattg	1905
Db	5074	GACCCTGAACCAAGGCGCCAGATAAGGAACAAGCTCGGGGCACTCTTCTGAAGGCCAAGCTGG	5015
Oy	1906	aggaagaaggagcagcgcgaacgcgtttggggagaagaatcctaaggaatccagactcattcctt	1965
Db	5014	AGGAAGAGGAGACAGCGCACCATTTTGGGAGAGATCTCAAGGATCCAAAGACTCTCAATTCCTT	4955
Oy	1966	tcctctggcccgagtaactttgtctctcccaagctttggggagaccccttccttggaacctta	2025
Db	4954	TCTCTTGCCCAAGTGAATTTGGTCTCTCCAGCTCTGGGGAGATCTTCTTGAACCTTA	4895
Oy	2026	ataaagaccccaactgagctctctctctctccatccctctccctctgacctctgaatttg	2085
Db	4894	ATAAAGACCCCACTGGAGTCTCTCTCTCCATCCCTCTCCCTGCGCTGCTTAATTG	4835
Oy	2086	ctgcacaggaatgtcactcccaaaccttactctgagctcaatlaataaataacagatttat	2145
Db	4834	CTGCACAGGATTTGTCACTCCAAACCTTACTGTGAGCTCATTAATAAAT---ACATTTAT	4779
Oy	2146	tttccagatttaa	2158
Db	4778	TTTTCCAGCTTATA	4766
RESULT 6			
G55504/c	G55504	463 bp	DNA
LOCUS	SHCC-100852	Human Homo sapiens	STS genomic, sequence tagged site
DEFINITION	G55504		
ACCESSION	G55504.1	GI:6120823	
VERSION			
KEYWORDS	STS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homidae; Homo.		
TITLE	1 (bases 1 to 463)		
JOURNAL	Myers, R.M.		
COMMENT	Human STS (1999)		
	Unpublished (1999)		

Initial incubation:	95 degrees C for 10 minutes
Denaturation:	94 degrees C for 30 seconds
Annealing:	60 degrees C for 30 seconds
Polymerization:	72 degrees C for 23 seconds
PCR Cycles:	30
Thermal Cycler:	perkin Elmer 9700
Protocol:	
Template:	25 ng
Primer:	each 1 uM
dNTPs:	each 200 uM
Amplifying Gold Polymerase:	0.07 units/uL
Total Vol:	5 uL

Buffer:	MgCl <sub>2</sub> :	2.5 mM
	KCl:	50 mM
	Tris-HCl:	10 mM
	pH:	8.3
BAC ends sequenced at TIGR from the RPC11 BAC library. Designed and developed at the Stanford Human Genome Center.		
FEATURES	Location/Qualifiers	
Source	1..463	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/map="11"	
	/clone_id="Human"	
STS	80..392	
primer_bind	80..102	
primer_bind	complement(370..392)	
BASE COUNT	96 a	116 c 128 g 123 t
ORIGIN		

Query Match	4.9%	Score 105	DB 13	Length 463
Best Local Similarity	82.8%	Pred. No. 4.5e-14		
Matches 120	Conservative 0	Mismatches 25	Indels 0	Gaps 0
Db 1502	ggtggcattgtgaaggagcgagatcttggaatcaacggaagaatgtgacagactac	1561		
Db 400	GGTGGCATTGTGAAAGGAGACAGATCATCGGAATCAACGGCAAGATTGTGACAGATTC	341		
0Y 1562	accctgagctgaagctcgacagctgcacctgcagaagcgcttgatccaggccgagctgagtc	1621		
Db 340	ACCCTGAGCTGAAGCTGACAGCTGCATTCGACGAAGGCTTGAAATCAGGCGGTAAGAAATA	281		
0Y 1622	gacctgtgtgtgtccgtctgcctccccc	1646		
Db 280	AGGCCCTCCCTCCCTTCTCCCTC	256		
RESULT 7				
LOCUS	I66494	7218 bp	DNA	28-DEC-1997
DEFINITION	Sequence 14 from patent US 5670367.			
ACCESSION	I66494			
VERSION	I66494.1	GI:2724471		
KEYWORDS	.			
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	1 (bases 1 to 7218)			
AUTHORS	Donner, F., Schellinger, F. and Falkner, F. Gunter.			
TITLE	Recombinant fowlpox virus			
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;			
FEATURES	Location/Qualifiers			
source	1..7218			
BASE COUNT	1944 a 1491 c 1486 g 1939 t 368 others			
ORIGIN				
Query Match	3.5%	Score 75.8	DB 5	Length 7218
Best Local Similarity	6.1%	Pred. No. 3e-07		
Matches 26	Conservative 240	Mismatches 157	Indels 0	Gaps 0
0Y 829	tcacctgtcgtcgtgagtgatgagatagaggccagatgctcgaagtcacatgctc	888		
Db 1478	TACCTATCATGCAAGTACTGTAAAGAGAGAGAAATTTGGTACRRRRRRRRRRRRR	1419		
0Y 889	gacctctcaacctgatacagaagagcgtgtaaatgtgctgaataatagccgcagctg	948		
Db 1418	RR	1359		
0Y 949	accatctccatgtgctgcagctgcgcggagcgttcattcatgacagaccggagcgctg	1008		

[illegible]

TITLE	COMMENT
<p>Direct Submission</p> <p>Submitted (04-JUN-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA</p> <p>On Jun 4, 1999 this sequence version replaced gi:4966400.</p> <p>All repeats were identified using RepeatMasker: Smit, A.F.A. &amp; Green, P. (1996-1997)</p> <p><a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a></p>	<p>Only 180385 base pairs from the middle of this clone are being submitted. The remainder overlaps either accession number AC005152 (WICGR project L335) or accession number AC005243 (WICGR project L343).</p>
FEATURES	
<p>source</p> <p>Location/Qualifiers</p> <p>1. 180385</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="84.E.24"</p> <p>/clone_11b="Alan Buckler -- per comm"</p> <p>/map="17"</p> <p>/chromosome="17"</p> <p>446..753</p> <p>/rpt_family="AluSc"</p> <p>2375..2419</p> <p>/rpt_family="(TTTG)n"</p> <p>complement(2985..3103)</p> <p>/rpt_family="L2"</p> <p>complement(3146..3352)</p> <p>/rpt_family="L2"</p> <p>complement(3531..3639)</p> <p>/rpt_family="L2"</p> <p>complement(3798..4394)</p> <p>/rpt_family="L2"</p> <p>5531..5631</p> <p>/rpt_family="MER81"</p> <p>5648..5674</p> <p>/rpt_family="(CAAAA)n"</p> <p>5812..6020</p> <p>/rpt_family="L1MC4"</p> <p>6180..6224</p> <p>/rpt_family="AT_rich"</p> <p>6303..6611</p> <p>/rpt_family="AluSp"</p> <p>7378..7618</p> <p>/rpt_family="MIR"</p> <p>9593..9623</p> <p>/rpt_family="(GGGA)n"</p> <p>complement(10666..10923)</p> <p>/rpt_family="AluSq"</p> <p>10958..11088</p> <p>/rpt_family="MIR"</p> <p>11752..11914</p> <p>/rpt_family="L2"</p> <p>complement(11915..12219)</p> <p>/rpt_family="AluSx"</p> <p>12220..12347</p> <p>/rpt_family="GA-rich"</p> <p>13110..13182</p> <p>/rpt_family="GA-rich"</p> <p>13526..13564</p> <p>/rpt_family="(TCC)n"</p> <p>13575..13660</p> <p>/rpt_family="(TA)n"</p> <p>complement(14380..14658)</p> <p>/rpt_family="AluJo"</p> <p>14742..14830</p> <p>/rpt_family="(TA)n"</p> <p>14860..14967</p> <p>/rpt_family="(TA)n"</p> <p>complement(14977..15122)</p> <p>/rpt_family="L2"</p> <p>complement(15292..15385)</p> <p>/rpt_family="L2"</p> <p>16672..16781</p>	







SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Homidae; Homo.  
TITLE 1 (bases 1 to 195385)  
JOURNAL Genome Therapeutics Corporation Sequencing Center: Human Genome  
REFERENCE Unpublished  
AUTHORS Smith, D.R.  
TITLE 2 (bases 1 to 195385)  
JOURNAL Direct Submission  
REFERENCE Submitted (08-DEC-1999) Genome Therapeutics Corporation, 100 Beaver  
AUTHORS Street, Waltham, MA 02453, USA  
COMMENT \* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 48 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1112: contig of 1112 bp in length  
\* 1113: gap of unknown length  
\* 2221: contig of 1109 bp in length  
\* 2222: gap of unknown length  
\* 3230: contig of 1009 bp in length  
\* 4630: gap of unknown length  
\* 3231: contig of 1400 bp in length  
\* 4631: gap of unknown length  
\* 5801: contig of 1170 bp in length  
\* 5801: gap of unknown length  
\* 7675: contig of 1875 bp in length  
\* 7676: gap of unknown length  
\* 9606: contig of 1931 bp in length  
\* 9607: gap of unknown length  
\* 10867: contig of 1261 bp in length  
\* 10868: gap of unknown length  
\* 12422: contig of 1555 bp in length  
\* 12423: gap of unknown length  
\* 13550: contig of 1128 bp in length  
\* 13551: gap of unknown length  
\* 14811: contig of 1261 bp in length  
\* 14812: gap of unknown length  
\* 15980: contig of 1169 bp in length  
\* 15981: gap of unknown length  
\* 18262: contig of 2282 bp in length  
\* 18263: gap of unknown length  
\* 19306: contig of 1044 bp in length  
\* 19307: gap of unknown length  
\* 21169: contig of 1863 bp in length  
\* 21170: gap of unknown length  
\* 22313: contig of 1144 bp in length  
\* 22314: gap of unknown length  
\* 23697: contig of 1384 bp in length  
\* 23698: gap of unknown length  
\* 25371: contig of 1674 bp in length  
\* 25372: gap of unknown length  
\* 26897: contig of 1526 bp in length  
\* 26898: gap of unknown length  
\* 28765: contig of 1868 bp in length  
\* 28766: gap of unknown length  
\* 32095: contig of 3330 bp in length  
\* 32096: gap of unknown length  
\* 34175: contig of 2080 bp in length  
\* 34176: gap of unknown length  
\* 36174: contig of 1999 bp in length  
\* 36175: gap of unknown length  
\* 38196: contig of 2022 bp in length  
\* 38197: gap of unknown length  
\* 40333: contig of 2137 bp in length  
\* 40334: gap of unknown length  
\* 42398: contig of 2065 bp in length

42399 44266: gap of unknown length  
\* 44267: gap of unknown length  
\* 47804: contig of 3558 bp in length  
\* 47805: gap of unknown length  
\* 51088: contig of 3284 bp in length  
\* 51089: gap of unknown length  
\* 55348: contig of 4260 bp in length  
\* 55349: gap of unknown length  
\* 58075: contig of 2727 bp in length  
\* 58076: gap of unknown length  
\* 61009: contig of 2934 bp in length  
\* 61010: gap of unknown length  
\* 64859: contig of 3850 bp in length  
\* 64860: gap of unknown length  
\* 68394: contig of 3535 bp in length  
\* 68395: gap of unknown length  
\* 74770: contig of 6376 bp in length  
\* 74771: gap of unknown length  
\* 78978: contig of 4208 bp in length  
\* 78979: gap of unknown length  
\* 85405: contig of 6427 bp in length  
\* 85406: gap of unknown length  
\* 90787: contig of 5382 bp in length  
\* 90788: gap of unknown length  
\* 97005: contig of 6218 bp in length  
\* 97006: gap of unknown length  
\* 102331: contig of 5326 bp in length  
\* 102332: gap of unknown length  
\* 109949: contig of 7618 bp in length  
\* 109950: gap of unknown length  
\* 118555: contig of 8606 bp in length  
\* 118556: gap of unknown length  
\* 126644: contig of 8089 bp in length  
\* 126645: gap of unknown length  
\* 136384: contig of 9740 bp in length  
\* 136385: gap of unknown length  
\* 149159: contig of 12775 bp in length  
\* 149160: gap of unknown length  
\* 162077: contig of 12918 bp in length  
\* 162078: gap of unknown length  
\* 181175: contig of 19098 bp in length  
\* 181176: gap of unknown length  
\* 195385: contig of 14210 bp in length.  
\* Location/Qualifiers  
1..195385  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-175M21"

BASE COUNT 56722 a 41286 c 40797 g 56532 t 48 others  
ORIGIN

Query Match 2.9%; Score 63.4; DB 43; Length 195385;  
Best Local Similarity 75.2%; Pred. No. 0.00028;  
Matches 79; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 854 agataggagaccagttgtcgaatgagtcgagtcgactctcaaccctgagtcacaagg 913  
|||||  
DB 12775 AGATAGGGGACCAAGATTGTCGAATGAGTCGAGTCGACTCTCAACCTGAGTCAACAAGG 12834  
QY 914 aggcgttaaatgtctgaaataagcgcgacgtcaccatcca 958  
||| | |||  
DB 12835 AGGTGAGATGTGGGGCTTCAACCTGTGGCCCTTGTATCTCCA 12879

RESULT 14  
AF148805/c  
LOCUS  
DEFINITION  
AF148805 28559 bp DNA VRL 02-AUG-1999  
Kaposi's sarcoma-associated herpesvirus ORF 68 gene, partial cds;  
and ORF 69, Kaposin, v-FLIP, v-cyclin, latent nuclear antigen, ORF  
K14, v-GPCR, putative phospholipase for myo-inositol synthase,  
and LAMP (LAMP) genes, complete cds.

ACCESSION	AF148805		
VERSION	AF148805.1	GI:5669888	
KEYWORDS	Kaposi's sarcoma-associated herpesvirus - Human herpesvirus 8.		
SOURCE	Kaposi's sarcoma-associated herpesvirus Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Rhadinovirus.		
ORGANISM	1 (bases 1 to 28559)		
REFERENCE	Glenn, M., Rainbow, L., Aurad, F., Davison, A. and Schulz, T.F.		
AUTHORS	Identification of a spliced gene from Kaposi's sarcoma-associated		
TITLE	herpesvirus encoding a protein with similarities to latent membrane		
JOURNAL	proteins 1 and 2a of Epstein-Barr virus		
MEDLINE	J. Virol. 73 (8), 6953-6963 (1999)		
REFERENCE	2 (bases 1 to 28559)		
AUTHORS	Glenn, M.A., Rainbow, L., Aurade, F., Davison, A.J. and Schulz, T.F.		
TITLE	Direct Submission		
JOURNAL	Submitted (15-Apr-1999) MRC Virology Unit, Church Street, Glasgow		
FEATURES	G11 5UR, UK		
Source	Location/Qualifiers		
gene	1..28559		
CDS	/organism="Kaposi's sarcoma-associated herpesvirus"		
gene	/isolate="GK18"		
CDS	/db_xref="taxon:37296"		
gene	/note="right end of the unique region plus 8 complete		
CDS	copies and 1 partial copy of the terminal repeat (TR)"		
gene	<1..2105		
CDS	/gene="ORF 68"		
gene	<1..1102		
CDS	/gene="ORF 68"		
gene	/function="involved in DNA packaging"		
CDS	/note="similar to HSV-1 UL32, HCMV UL52 and EBV BFLF1"		
gene	/codon_start=2		
CDS	/evidence-not_experimental		
gene	/protein_id="AAD46496.1"		
CDS	/db_xref="GI:5669889"		
gene	/translation="IMHLPPEEATYGLFGGRMGIDLOLHFPVOKCFKTAEKI		
CDS	LGISNLOFLKSEFIRGMLTGITTCNCFKRTSNPRIDKEATGPTCCOITDTPPAS		
gene	GPELAKATFCGASRPKPSLIPALIDIMSTSSLEDEPRPRLNSDMSLKVASH		
CDS	DPEFSPLOADTSQPCLMHPPLGLRYNGTASVCLCECLAAHEAKALQTLQCEV		
gene	MGHLENNVKLVRIAEVLNDPEFAMPYVSDPLIRELRCTQEIHKHLEFCALINA		
CDS	KVSEEDVFLRLPREDEYKKLRASAAAGOLDANLTFDCEVQTVLEFLFKGNARVGK		
gene	TTSLDIIRETLAQLRHRLDLAHSQTSHTYA"		
CDS	1135..2105		
gene	/gene="ORF 69"		
CDS	1135..2043		
gene	/gene="ORF 69"		
CDS	/note="similar to HSV-1 UL31, HCMV UL53 and EBV BFLF2"		
gene	/codon_start=1		
CDS	/evidence-not_experimental		
gene	/protein_id="AAD46497.1"		
CDS	/db_xref="GI:5669890"		
gene	/translation="MPKSVSSHISLATSTGSGRPDIRRCLSSRLSVPGARSASVS		
CDS	SKHRNGLRKPTSDKVFESILSHRELGDPLREMPICSTVWLPLDLISVAGRC		
gene	VLSRFGSSNMNGOCALCPSTENPTVAGSGSPOTMGDALKKNNELSVALARINHA		
CDS	DKVIOHKTFYLSLSHSDVVRQSTLQGLPLANVLKTFGHDPLPITTTNGMLTWC		
gene	ILFTRALHIGETALRLMLMDNLPNYKISADCCROSSYVVFVPTHTDTSIAVQHTIC		
CDS	EVAVALDCTDEKRDIDOKGTALVNAL"		
gene	2100..2105		
CDS	/gene="ORF 69"		
gene	/note="possible signal for ORFs 68 and 69"		
CDS	/evidence-not_experimental		
gene	/complement(2126)		
CDS	/gene="ORF K12"		
gene	/evidence-experimental		
CDS	/complement(2126..2824)		
gene	/gene="ORF K12"		
CDS	/product="T0.7"		
gene	/evidence-experimental		
CDS	/complement(2126..2824)		
gene	/gene="ORF K12"		
CDS	/complement(2139..2144)		
gene	/gene="ORF K12"		
CDS	/evidence-experimental		
CDS	complement(2616..2798)		
gene	/gene="ORF K12"		
CDS	/note="transferring gene"		
gene	/codon_start=1		
CDS	/evidence-experimental		
gene	/product="Kaposin"		
CDS	/protein_id="AAD46499.1"		
gene	/db_xref="GI:5669892"		
CDS	/translation="MDKGLTVFAVAVPVDLLNGWRMLGAIPLVLCILASVYPPSG		
gene	ORGPVAFRTVATGANH"		
CDS	complement(2799..2824)		
gene	/gene="ORF K12"		
CDS	/evidence-experimental		
gene	/rpt_type=direct		
CDS	2820..3081		
gene	/rpt_type=direct		
CDS	/rpt_unit=2820..2842		
gene	3082..3436		
CDS	/rpt_type=direct		
gene	/rpt_unit=3084..3106		
CDS	complement(6928..6933)		
gene	/gene="ORF 71"		
CDS	/note="possible signal for ORF 71"		
gene	/evidence-not_experimental		
CDS	complement(6928..7550)		
gene	/gene="ORF 71"		
CDS	complement(6984..7550)		
gene	/gene="ORF 71"		
CDS	/function="inhibitor of apoptosis"		
gene	/note="contains two death effector domains"		
CDS	/codon_start=1		
gene	/evidence-experimental		
CDS	/product="v-Flip"		
gene	/protein_id="AAD46498.1"		
CDS	/db_xref="GI:5669891"		
gene	/translation="MATYEVCEYARKLGTDDREYVFLNWFIPQPTLAIGALRA		
CDS	LKEGRLEPLLAECLEFRAGRDRLDLHLDPRELEAGTMSYSPYQTYLAVD		
gene	GELCARLRSILFLSKDTIGSRSTQOTFLHWYTCMENLDLGPDDVDALMSLSLSR		
CDS	VDLQOVOTLWGLHLSGSHQHRHTP"		
gene	complement(7601..7606)		
CDS	/gene="ORF 72"		
gene	/note="possible signal for ORFs 72 and 73"		
CDS	/evidence-not_experimental		
gene	complement(7601..8406)		
CDS	/gene="ORF 72"		
gene	complement(7601..12037)		
CDS	/gene="ORF 73"		
gene	complement(7633..8406)		
CDS	/gene="ORF 72"		
gene	/note="cyclin D homologue"		
CDS	/codon_start=1		
gene	/evidence-experimental		
CDS	/product="v-cyclin"		
gene	/protein_id="AAD46500.1"		
CDS	/db_xref="GI:5669893"		
gene	/translation="WATANNPPSGLDLPTCEDRIYVILIEPRELTSDSVFTGQO		
CDS	SLTSHMRKLGTWPFVSCQENLEBNVVALMLDLRLLLKQVSKERFQTSACIL		
gene	VASKRSRLPSTSSICYAADSFSROGLIDQEKILKAMRTFAVATVTSLL		
CDS	KLGGSHLDPEWHHEVNTLTKALVDPTGSLPASIIAAGCALLVPANVLPDTHSG		
gene	GVVQGLASITICDVSVAQVILTSVSDPDLRLDSY"		
CDS	complement(8648..12037)		
gene	/gene="ORF 73"		
CDS	/note="LNA: component of latency-associated nuclear		
gene	antigen: LANA"		
CDS	/codon_start=1		
gene	/evidence-experimental		
CDS	/product="latent nuclear antigen"		
gene	/protein_id="AAD46501.1"		
CDS	/db_xref="GI:5669894"		
gene	/translation="WAPGMRLRSRGTGAPLTRSCRRNRSPERCIDLGLDLPQR		
CDS	RKHVADSVGREGCPHTLPITGSPVTFGAPVSPPLPVPAPAPAPPPA		
gene	LLPVTSSSP		
CDS	MRPPSQQTTPPHSPPTPPPPPPSSPDSLAPSTLSLRKRRRLSSPPGSPSTLPNICO		

RESULT 15	AC016961	230309 bp	DNA	HTG	19-FEB-2000
LOCUS	AC016961				
DEFINITION	Homo sapiens clone RP11-394J21, WORKING DRAFT SEQUENCE, 53				
ACCESSION	AC016961				
VERSION	AC016961.7	GI:7007897			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE 1	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE 2	(bases 1 to 230309)				
AUTHORS	Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K., Bodoir,B., Bouch,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C., Burdett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C., David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N., Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D., Forcum-Tansley,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L., Guevara-M., Harris,K., Hernandez,J., Hodgson,A., Hoques,M., Holloway,C., Hosack,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondelowski,N., Kong,Y., Kovar,C., Leal,B., Li,Z., Lichteberg,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mel,G., Morgan,M., Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugeng,R., Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wabnah,M., Wellington,S., Weinstein,G., Weinstein,I.R., Williamson,A., Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.				
TITLE	Direct Submission				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 230309)				
AUTHORS	Worley,K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-DEC-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
COMMENT	On Feb 19, 2000 this sequence version replaced gi:6715662.				

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: HMV

Center clone name: RP11-394J21

----- Summary Statistics

Sequencing vector: M13, L08821

Chemistry: Dye-Primer Bodipy: 5% of reads

Chemistry: Dye-terminator Big Dye: 95% of reads

Assembly program: Phrap; version 0.980611

Consensus quality: 180544 bases at least Q40

Consensus quality: 195032 bases at least Q30

Consensus quality: 202601 bases at least Q20

Estimated insert size: 155606; agarose-fp estimation

Estimated insert size: 214941; sum-of-ctnigs estimation

Quality coverage: 4.2x in Q20 bases; agarose-fp estimation

Quality coverage: 3.1x in Q20 bases; sum-of-ctnigs estimation

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 53 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

1 753: contig of 753 bp in length





**THIS PAGE BLANK (USPTO)**